

Dispatches from the front line

The rich world's corridors of power echo with talk about the new 'war on terrorism'. Meanwhile, most developing countries are losing another war that has already inflicted millions of casualties.

The threat was first identified some two decades ago. Since then, more 20 million people have been killed, and unless immediate offensive strikes are made, intelligence sources predict the slaughter of another 68 million within 20 years. That's around 9,000 victims per day, all claimed by HIV. No continent will be spared: the war against AIDS — the hackneyed metaphor is in this case an understatement — is truly a world war.

Virologists have been mobilized, but with hopes of an effective vaccine still years down the road, the immediate imperative has less to do with science and more to do with politics and hard cash. We already have the weapons — prevention schemes and anti-retroviral drugs — to prevent some 29 million new infections by 2010 (J. Stover *et al. Lancet* **360**, 73–77; 2002). What's missing is the political will and the money to deploy them quickly where they are most needed: in the developing countries where 95% of cases occur.

The international community is at least talking a good fight, having pledged in June 2001 to “reduce by 2005 HIV prevalence among young men and women aged 15 to 24 in the most affected countries by 25%”. To achieve this, it promised an annual spend of US\$7–10 billion. But so far just \$2 billion has actually materialized for the Global Fund to Fight AIDS, Tuberculosis and Malaria. And while a handful of developing countries have made AIDS their top priority, many continue to deny and stigmatize the disease, failing to prioritize the investments in health infrastructure that will be needed if any injection of new donor funds is to be effective.

Some economists and politicians still argue for putting all available cash into preventing new cases, arguing that supplying anti-retroviral drugs to much of the developing world is not cost-effective. But this flies in the face of the evidence and thankfully is becoming a minority view. In reality, the human, social and economic benefits of widening

access to drugs are obvious. People live longer, so they can work to support their families, and fewer children are orphaned — not to mention the fact that anti-retrovirals could prevent the transmission of HIV to some three million babies each year.

Moreover, the availability of drug treatment seems to create a virtuous circle, reducing the spread of HIV. People are more motivated to take diagnostic tests, and the vastly reduced viral load in most treated individuals reduces the risk of them infecting others.

If we can get cheap Coca-Cola to every corner of Africa, there is no reason why we can't deliver anti-retroviral drugs there too. Three years ago this proposal would have seemed outrageous. Annual drug treatment for one person cost between \$10,000 and \$15,000, and the logistics were a nightmare. But the cost of treatment has fallen to as low as \$300, and the aid agency Médecins sans Frontières reckons it can be brought down to \$50. In part, this is due to the Joint United Nations Programme on HIV/AIDS and the World Health Organization brokering price deals with drugs companies. A more successful strategy comes from countries such as Brazil, Thailand and Cameroon, which have used the threat of compulsory licences to produce generic alternatives as a tactic to negotiate tough deals with the big drugs firms.

Brazil alone has treated 115,000 patients. Its genuine and sustained commitment to prevention and to purchasing and delivering drugs means that a national epidemic that was predicted to reach 1.2 million cases this year instead stands at 600,000, and AIDS-related deaths have been slashed by half. What's more, the policy has saved government money by cutting the costs of caring for AIDS patients with opportunistic infections.

This evidence speaks for itself. The time for lukewarm promises and glacial progress is over. The world's rich nations must start filling the Global Fund's coffers, so it can fight the war on AIDS in earnest. ■

All together now...

Britain's chief scientific adviser wants to revamp the government's fragmented approach to science. Let's hope he succeeds.

The job of chief scientific adviser to the British government has long entailed large doses of frustration — the position has come with no direct authority over the civil servants who control the research programmes of individual government departments. “I do not have an authority, I have merely an advisory capacity,” the previous incumbent, Robert May, told the House of Commons Science and Technology Committee in February 2000.

As a result, insular attitudes frequently got in the way of a coherent approach to research and the provision of scientific advice — dynamics that contributed to the initially sluggish response to the outbreak of foot-and-mouth disease that last year devastated British agriculture.

Chemist David King, who took over as chief scientific adviser at the end of 2000, was given a baptism of fire by foot-and-mouth. But the crisis made him focus him on the need to change the culture that dissuaded the now-defunct Ministry of Agriculture, Fisheries and Food from seeking the independent advice that might have prevented matters getting so badly out of hand. When interviewed

last summer, King talked of the need to “parachute in” experts from academia and industry to replace the career civil servants who have long held sway over science within individual ministries (see *Nature* **412**, 472–473; 2001).

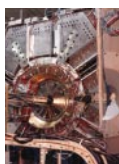
It is encouraging, therefore, to see that King's fingerprints are on the government's new science strategy, unveiled last week (see page 472). Every department with an appreciable science budget must now appoint its own chief scientific adviser who will have direct access to their cabinet minister. King will presumably corral this group of experts to exert the coordinating influence that his predecessors were unable to muster.

No one should underestimate the difficulty of challenging the prevailing culture of Britain's civil service. But encouragingly, King seems to have the backing of Britain's two most powerful politicians, Prime Minister Tony Blair and Chancellor of the Exchequer Gordon Brown. He also deserves the support of the wider British scientific community, which stands to gain a stronger voice within government. ■



-82.24
7702.34
-17.81
-81.15
+26-477

Money meltdown
Stock-market crashes
blitz universities'
share donations
p468



Matter of fact
Physicists give a
thumbs-up to the
standard model
p469



Protest vote
Objections prompt
restriction of
stem-cell patent
p470



Sushi sequence
Researchers get
their hands on
pufferfish genome
p472

Row over neutron source heats up as Germany's advisers cry foul

Quirin Schiermeier, Munich

When Germany's science council, the Wissenschaftsrat, declined to recommend a major new neutron source for funding, it said it was acting on the advice of an international committee of physicists. Now members of that expert group claim that their views were distorted, and that the group's chairman inserted remarks critical of the neutron project without their blessing.

The council decided on 12 July that German designs for the European Spallation Source (ESS), a 1.4-billion-euro (US\$1.4-billion) neutron facility that will be used to probe molecular structures, did not deserve funding. The move was based on a report by a 15-member panel chaired by Hans Spiess, a director at the Max Planck Institute for Polymer Research in Mainz. The council concluded that alternative and cheaper technologies were likely to become available by 2011, the earliest time at which the ESS could become operational (see *Nature* **418**, 262; 2002).

But on 19 July, seven members of the group wrote to the council's president, Karl Max Einhäupl, claiming that the report "reflected neither the consensus view of the subcommittee nor the opinion of most

members" (see *Nature* **418**, 367; 2002). Einhäupl has since said that the group was given three weeks to comment and amend the report in April, and that all but one endorsed it (see Correspondence, page 479). But panel members say that Spiess did not include their comments in the final report.

"The final draft of our report, which I received four days before it was adopted, bore no resemblance to what we had discussed," says Thomas Mason, director of the Spallation Neutron Source currently under construction at Oak Ridge National Laboratory in Tennessee. "There was also no indication of a deadline for making objections, and by the time I sent a note describing my concerns it was already too late."

Other signatories of the letter report similar experiences. "The final draft of our report included critical remarks about the ESS that we had never discussed," says Erich Sackmann, a biophysicist at the Technical University of Munich. "Then suddenly everything went very quickly, and the report was endorsed without further discussion, not to mention a final meeting or vote."

Sackmann and Mason allege that addi-



Model idea: but the report concluded that cheaper alternatives to the source were imminent.

tions to the report, concluding that developments in nuclear-magnetic-resonance imaging and synchrotron technology were likely to replace neutrons, and that a 'neutron drought' alone did not justify major investment in a new facility, were made by Spiess.

Spiess, who has previously criticized the neutron community for its narrow scientific horizons, emphatically rejects having weakened the case for the ESS in the report. "After we disseminated the report in April we received very positive comments," he says, adding that only one person objected. ■

► www.wissenschaftsrat.de/texte/5373-02.pdf

Accused obesity researcher returns to the French fold

Declan Butler, Paris

Bernard Bihain, the obesity researcher who left French research in 1998 amid allegations of scientific misconduct, is to take up a post with his former employers INSERM, France's national biomedical research agency.

Bihain's claim to have identified and cloned a molecule involved in fat degradation was queried in 1997 by members of his lab at the University of Rennes 1. An internal inquiry commissioned by the French research ministry concluded that the testimony of seven of the whistleblowers raised doubts about certain results produced by the lab. But no action was taken, and plans for a second, international inquiry were shelved after Bihain closed his

lab in August 1998 (see *Nature* **395**, 829; 1998).

After working in the United States for the French biotech company Genset and subsequently for ValiGen, a genomics firm, Bihain returned to France and applied for a new position at INSERM. He is expected to work at the agency's centre for clinical investigation in Nancy, leading a group researching into anti-obesity drugs.

Christian Brechot, director-general of



Bernard Bihain is set to take up a new post in Nancy.

INSERM, says that Bihain is entitled to a permanent INSERM post under French law. He argues that his decision to hire Bihain was based on the researcher's overall scientific record and that the misconduct allegations are a separate issue for the science ministry. But he says he did ask an international team of experts to review a report commissioned from Bihain detailing his past research and future plans. "The opinion was that although some of his hypotheses have not been confirmed, his contribution was of value," Brechot says.

Bihain, who always denied the allegations, says the charges have never been proved. "I would like to be shown one published scientific fact that is false," he says. ■

Falling share price stymies gift to university

Rex Dalton, San Diego

Researchers at the University of Colorado thought they had tapped a rich vein of gold last year when a Silicon Valley entrepreneur pledged \$250 million in shares for studies of cognitive disorders. But recent market falls have wiped 90% off the value of the shares, delaying some Colorado research plans and sending a warning to other institutions that accept such gifts.

The shares in BEA Systems, a business-software company based in San Jose, California, were due to be donated by the company's co-founder William Coleman and his wife Claudia. The couple planned to give the university around \$50 million a year for five years to set up an endowment fund for research into areas such as developmental disabilities, Alzheimer's disease and strokes.

They gave an initial \$15 million in cash early last year to set up the Coleman Institute for Cognitive Disabilities, based at the uni-



A funding pledge to back research at the Coleman Institute for Cognitive Disabilities (above) has been held up by the volatile stock market (below).

versity's Boulder campus, and to set up 15 research projects, including studies of stem cells and nerve regeneration. The rest of the year's instalment of \$40 million in stock was due to be paid last November. But the couple decided to delay the gift until 2006 when BEA's share price fell from more than \$70 to less than \$20 per share during 2001. With the

price currently at just above \$5, this autumn's \$50-million donation may also be delayed.

The Colemans have donated a further \$1.6 million to cover the income that the university would have expected to receive from its endowment, ensuring that the research projects continue. But without the security of an endowment, the university is wary of expanding too quickly and some projects have not progressed as rapidly as expected.

Shares are a common way of donating to universities, but most institutions have a policy of selling them immediately. Over the past two years, for example, the University of California, San Diego, has received a \$20-million donation from computer scientist John Moores, chairman of the board at San Diego-based software company Peregrine Systems, and his wife Rebecca. The university sold the shares immediately, and ploughed the proceeds into a new centre for cancer research and patient care, due to open in 2004. The stock, which peaked at over \$60 per share in early 2000, now stands at around 27 cents.

Patrick Mulvey, vice-president for development at the M. D. Anderson Cancer Center in Houston, Texas, says his centre was fortunate to lose only \$600,000 in pledges when Enron went bankrupt in December. The Houston-based energy company had previously given \$1.6 million to the centre. Stock market volatility is making donors wary of gifts of shares, Mulvey adds. ■

INDU -82.24	VOLU 2.434
INDP 7702.34	UVOL 469.0
UTIL -17.81	DVOL 1.931
TRAN -81.15	TRIN .87

Asteroid rating system makes an impact on the media

Jonathan Knight, San Francisco

A new scale for assessing the risk of asteroid impacts was thrust into the limelight last week. The Palermo Technical Impact Hazard Scale is not intended for public consumption, but it sparked media interest when a two-kilometre-wide asteroid produced the scale's highest rating yet.

Since 1999, nearby asteroids have been rated on the Torino scale, in which zero represents no risk and 10 means that a globally devastating collision is almost certain. The scale was devised to communicate relative risks of asteroids to the public after media reports in 1998 that an asteroid might end civilization in 2028. Further calculations showed the asteroid to be harmless.

The Torino scale is not sensitive enough to help astronomers judge which of the dozens of nearby objects merit attention. "We wanted a way to look at a potential impact and to ask 'is it worth some big telescope time?'," says Steven Chesley, an astronomer at NASA's Jet Propulsion

Laboratory in Pasadena, California, who led the development of the new scale.

For each asteroid, the Palermo scale rates the chance of its impact relative to the frequency with which objects of a similar size collide with the Earth. A rating of zero means that an impact is no more likely than a similar but as yet unidentified object

striking the Earth before the asteroid being studied is scheduled to arrive. The value is expressed as a logarithm, so -2 implies that a collision is 1% as likely as the 'background' probability, and a rating of 2 means that an impact is 100 times more likely.

Initial measurements of the trajectory of the newly spotted asteroid — dubbed 2002 NT7 — indicated that it had roughly a 1-in-100,000 chance of striking the Earth in 2019. The asteroid is two kilometres across, and an object of this size is predicted to hit the Earth about every million years.

The figures gave 2002 NT7 a positive Palermo value — the first since the scale was adopted — and ensured widespread media coverage. "I was surprised to see a positive value within the first year of operation," says Chesley.

Further measurements have now reduced 2002 NT7's Palermo value to below zero and ruled out the chances of a collision in 2019, although the possibility of it striking the Earth in 2053 remains open. ■

♦ <http://neo.jpl.nasa.gov/risk>



Model success is not the end of the matter

Geoff Brumfiel, Washington

The standard model of high-energy physics received a ringing endorsement last week, as researchers revealed the most accurate measurements so far taken of an effect that may explain the surplus of matter in the Universe.

The results, presented at the International Conference on High Energy Physics in Amsterdam by teams from the United States and Japan, focus on a fundamental effect that some think could explain why we exist.

Physicists believe that at the time of the Big Bang, the Universe contained equal amounts of matter and antimatter. But these should have annihilated each other, leaving no matter to create galaxies, stars — or us. One explanation for the glut of matter in the Universe is an effect called charge-parity (CP) violation, which causes matter and antimatter to behave in slightly different ways.

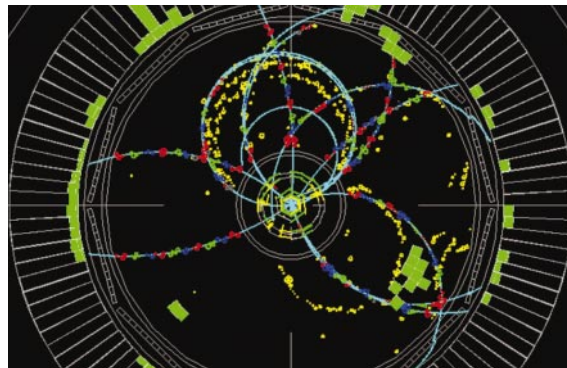
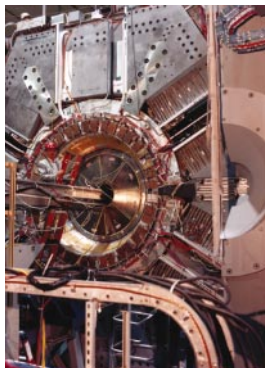
Two groups — the Belle collaboration based at the High Energy Accelerator Research Organization in Tsukuba, Japan, and BaBar at the Stanford Linear Accelerator Center in the United States — have measured the decay of particles called B mesons and their antiparticle equivalent, anti-B mesons. Both teams detected a slight difference in the decay rates of the particles, which is interpreted as evidence for CP violation.

The measurements, which are around twice as accurate as those previously taken (see *Nature* **412**, 105; 2001), agree well with each other and also with the standard model's predictions. "The agreement is as perfect as it could be," says Stewart Smith, a physicist at Princeton University and spokesman for the BaBar collaboration.

Despite this support for the standard model, current measures of CP violation show that the effect is not strong enough to account for the excess of matter. "The measurement fails by a factor of ten billion or so," Smith says. "There's something out there that we don't know about yet. These latest measurements don't even give a hint."

Some physicists believe that a possible explanation could be the existence of super-heavy particles that are not predicted by the standard model. The presence of these particles in the early Universe would have influenced the balance between matter and antimatter, tipping the scales in the favour of matter.

Delegates in Amsterdam were due to hear tentative evidence for super-heavy particles this week. The g-2 experiment at the Brookhaven National Laboratory in New York state measures a property of particles called muons that should deviate slightly



Data on particle decays (right) obtained using BaBar's equipment vindicate physics' standard model.

from the predictions of the standard model if super-heavy particles do indeed exist. The latest results from Brookhaven hint at this deviation, but physicists will treat the results

with care after a similar announcement last year turned out to be based on a flawed calculation (see *Nature* **410**, 291; 2001 and *Nature* **415**, 6; 2002). ■

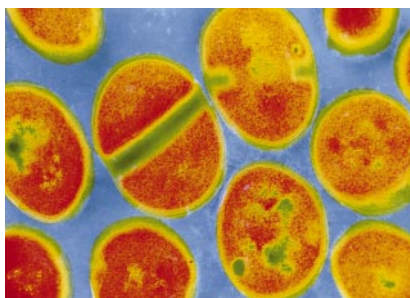
'Superbug' hurdles key drug barrier

Helen Pearson

Microbiologists have expressed fears that hospitals are ill-equipped to cope with 'superbugs', after a bacterium resistant to one of the last lines of antibiotic defence was identified in the United States.

The US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, reported last month that a strain of vancomycin-resistant *Staphylococcus aureus* had been isolated in June from the foot ulcer and catheter of a patient in Michigan. *S. aureus* often causes severe infections in patients' wounds, and outbreaks can lead to whole hospital wards being shut down.

Vancomycin became a vital weapon against *S. aureus* when strains resistant to the preceding antibiotic of choice, methicillin, emerged in the 1980s. But since 1987, when relatively harmless gut bacteria called enterococci evolved resistance to vancomycin, microbiologists have been warning that these genes could be transferred to *S. aureus*.



Tiny terror: outbreaks of *Staphylococcus aureus* could pose a severe threat to safety in hospitals.

DNA sequence analysis revealed that the Michigan *S. aureus* strain did acquire its resistance genes from an enterococcus. A vancomycin-resistant *Enterococcus faecalis* isolated from the same patient could explain the jump, suggests the CDC's Fred Tenover, who led the analysis.

The outbreak is probably the first of many, as vancomycin-resistant enterococci are increasing in prevalence — almost a third of enterococci in US intensive-care units are now resistant to the antibiotic. Careful monitoring and rigorous control of outbreaks are the best response, says Gary French, a microbiologist at Guy's & St Thomas' Hospital in London.

Drug-resistant bacteria are usually detected in the lab by their ability to grow near paper discs soaked in antibiotic. But this method can fail with vancomycin, as it does not diffuse well, giving unusual patterns of bacterial growth that lab staff may not recognize. A back-up test, in which bacteria are grown on plates soaked in vancomycin, may not be used because of the costs involved.

If there is a repeat of the Michigan incident, French fears that some hospitals lack the staff and isolation rooms to control the infection. "We know what we ought to do but I'm concerned that there aren't enough resources to do it," he says.

Vancomycin resistance is not as worrying as it would have been before the introduction of new antibiotics called linezolid and quinupristin/dalfopristin in 1999–2000, when vancomycin was often the last line of defence. But clinical biologists are keen to limit the use of these newer drugs, as some *S. aureus* strains are already resistant to them. ■

Biomedical institute suffers growing pains

Erika Check, Washington

With a new director on the way and its first major round of grants set to be awarded later this year, the US National Institute of Biomedical Imaging and Bioengineering (NIBIB) is ready to roll.

But this latest member of the National Institutes of Health (NIH) is already ruffling its siblings' feathers, as researchers watch existing programmes move to the new body — and wonder if their project will be next.

The NIBIB caused unease even before it came into existence in 2000. Many NIH staff, including former director Harold Varmus, opposed it, saying that it would disrupt existing research and increase NIH bureaucracy.

And the arrival this May of the NIH's new director, Elias Zerhouni, a radiologist by training who backed the NIBIB's creation before he was appointed, caused speculation that it could be favoured above other institutes. "There is a lot of anxiety about the new institute among NIH staff and NIH-funded researchers," says one senior NIH official.

Few NIH researchers are willing to criticize the NIBIB in public, but many are voicing their fears in private. Doubts centre on the group of experts charged by Congress with setting the institute's budget and research agenda. Six of the nine-member group were representatives of two bodies that backed the NIBIB's creation — the Academy of



Elias Zerhouni says tensions over the biomedical imaging and bioengineering institute will ease.

Radiology Research and the American Institute for Medical and Biological Engineering. The remaining three were NIH staff.

Of the \$112-million budget set for the NIBIB this year, \$67 million came from funds transferred from other institutes. But the group has recommended that a further \$150 million be transferred in the next financial year. "The money has to come from some-

where, and nobody wants to see a reduction in their area," says another NIH official.

The process by which grants are chosen for transfer is also raising hackles. The first round was picked by NIH officials, but the second list was drawn up earlier this year by the working group. Some NIH staff say many programmes that are disease-specific should have stayed in other institutes.

"A lot of the transfers didn't make sense," agrees Mike Marron, director of the Division of Biomedical Technology at the NIH's National Center for Research Resources. "The process certainly was not transparent."

Donna Dean, acting director of the institute, says that the new body is focusing on previously undersupported areas. Sensors to detect weapons of mass destruction have, for example, become a hot topic since 11 September. "Sensors seemed to be an area that no one was taking strong ownership of," she says. "We thought this is an area we can take on."

In the long run, the NIBIB's fate is in the hands of new director Roderic Pettigrew, a medical-imaging expert from the Emory University School of Medicine in Atlanta, Georgia, who arrives in September. And Zerhouni is confident that the institute will work through its growing pains. "This sense of tension is a natural epiphenomenon of the creation of any new initiative," he says. "I don't see that it will have lasting effects." ■

K. LAMBERT/AP

Opponents of stem-cell patent win restrictions

Alison Abbott and Oliver Schmidt, Munich

A controversial patent covering technologies for purifying human and animal stem cells has been restricted in scope by the European Patent Office (EPO).

In a ruling issued on 24 July, the EPO's oppositions division narrowed the patent held by the University of Edinburgh's Centre for Genome Research, invalidating all claims involving animal and human embryonic stem cells.

The patent, which details techniques for isolating and propagating genetically engineered adult or embryonic stem cells, generated widespread complaints when it was granted in December 1999.

Starting in March 2000, 14 parties, including Greenpeace and the German, Italian and Dutch governments, filed objections. Most opposed the patent because it involved techniques for working with human embryonic cells (see *Nature* 404, 3–4; 2000). The DFG, Germany's main research funding agency, objected on technical grounds, claiming that the University of Edinburgh had not revealed enough technical

information about the methods involved.

Shortly after the objections were filed, the university withdrew parts of the patent that covered technologies that could be used to alter the composition of the human germ line. But the EPO's ruling on the formal objections has gone further, leaving only claims covering adult stem cells intact.

The patent office says that its decision was partly based on ethical grounds — uses of human embryos are excluded from patentability according to EPO rules — and partly due to the patent's failure to disclose sufficient information for the techniques to be repeated by stem-cell experts.

Oliver Brüstle, a stem-cell researcher at the University of Bonn, fears that the ruling will send a negative signal to researchers. "It could mean that it will be difficult for researchers to get protection for intellectual property on human embryonic stem cells in the future," he says.

But Bernard Huber, a lawyer for the DFG, points out that the EPO's decision applies only to this particular case, and so cannot be generalized. ■

C. LEHSTEN/GREENPEACE/AP



Live issue: since it was granted in 1999, the 'Edinburgh patent' has faced stiff opposition.

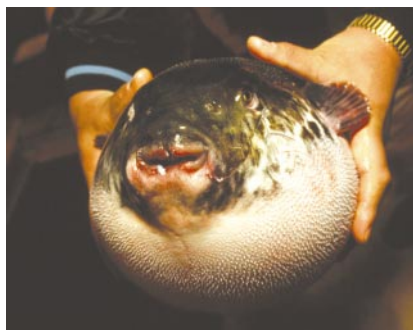
Genome projects net pufferfish and zebrafish sequences

London Gene hunters and developmental biologists are preparing for a feast in the form of two fresh fish genomes, one complete and one in a useful draft form.

The genome of the pufferfish (*Fugu rubripes*) was published online last week by an international consortium (S. Aparicio *et al.* *Science* 10.1126/science.1072104). At 365 million base pairs, it lacks much of the 'junk' DNA found within the human genome, fitting roughly the same number of genes into a genome about one-tenth of the size. This should make it easier to identify genes.

A preliminary assembly of the zebrafish (*Danio rerio*) genome was also released last week by another consortium led from the Sanger Institute at Hinxton, near Cambridge, UK. The data will help researchers to make sense of the thousands of zebrafish mutants that are studied around the world. Zebrafish embryos are transparent and develop quickly, making the zebrafish a favourite model organism in vertebrate developmental biology. Although the sequence is an early draft, the team says it is releasing the data "because there is an enormous amount of useful sequence and this outweighs the problems in the assembly".

♦ www.ensembl.org/Danio_rerio



Inflated importance: the pufferfish has the same number of genes as a human, but less junk DNA.

Senate committee plans mixed budget news for NSF

Washington A US Senate committee has recommended increasing funding for the National Science Foundation (NSF) by 11.8% next year — but wants to slash the foundation's budget for big projects by almost a half.

The proposed budget released by the Committee on Appropriations late last week calls for a \$645 million increase in funding, almost double the rise requested by President George Bush (see *Nature* 415, 565; 2002). But it also recommended cutting 43% from the

NSF's large-projects fund by delaying a major environmental monitoring project, known as the National Ecological Observatory Network, and withholding \$20 million from EarthScope, a series of large projects in the Earth sciences.

In part, the cut reflects concerns in Congress about the NSF's management of large projects. The EarthScope money, for example, will be restored when the NSF dedicates a staff member to manage the construction budget for such projects.

The House of Representatives is currently drafting its own version of the spending bill, which must be reconciled with the Senate bill and signed by President Bush later this year.

Fears of data misuse spur publishing debate

Washington The US National Academy of Sciences (NAS) plans to hold a meeting of publishers, scientists and national-security experts to examine how scientific journals should handle sensitive data that could potentially be used by terrorists.

The meeting, which is expected to take place in the autumn, is being called in response to a letter sent by Ron Atlas, president of the American Society for Microbiology, to the NAS president Bruce Alberts on 22 July. Atlas pointed out in the letter that the society has received many requests from researchers who wish to withhold data from papers that they submit to its journals (see *Nature* 415, 821; 2002). He wrote: "The purpose ... would be to initiate discussion about the possible development and eventual adoption of common publication policies."

"There's been a lot of talk about these issues, but we don't know how real the problem actually is," says NAS spokesman Bill Kearney. "We hope this meeting will help us to find out."

Moon samples safe as FBI ends students' rock folly

Washington Three NASA summer interns have tried to supplement their stipends by stealing Moon rocks and selling them over the Internet. The gang, plus one accomplice, has been charged with taking a safe containing 130 grams of Moon samples from NASA's Johnson Space Center in Houston, Texas. They drove the safe to Tampa, Florida, and attempted to auction off the contents on a Belgian website.

NASA officials were unaware that their rocks had gone missing until they received a tip from a Belgian mineralogist. In an undercover operation, FBI officials negotiated a deal with the interns, then arrested them when they showed up to sell the specimens.

Election seeds end to transgenic crop ban

Wellington A political crisis over genetically modified (GM) crops in New Zealand seems to have been averted. In the run-up to the general election on 27 July, the country's current moratorium on the release of GM organisms became a prominent campaign issue.

The centre-left Labour government of Prime Minister Helen Clark had signalled that it would not extend the moratorium, which is set to expire in October 2003. But the Green Party, which many political pundits believed would end up holding the balance of power after the election, was determined that the moratorium should be retained.

The row became a hot issue with the release of a book called *Seeds of Distrust*, which alleged from leaked documents that Clark's government had concealed an incident in which maize seeds contaminated with GM material were planted over an area of 178 hectares.

In the event, however, last week's poll gave Clark's party enough seats to form a minority government without having to rely exclusively on Green support.

Britain unveils strategy for spending new science cash

London First British science was given a major injection of funds, now it has a new strategy. On 23 July, the government released *Investing in Innovation*, a document that details changes that will accompany the £1.25-billion (US\$1.95-billion) increase in funding announced earlier in the month (see *Nature* 418, 261; 2002).

In return for the extra money, government departments will have to prepare detailed budgets for science and innovation. All government-funded research will also be subjected to scrutiny by outside experts — the details and costs of which will be decided in the autumn.

Under the new strategy, David King, the government's chief scientific adviser, should get more authority to coordinate the science programmes of various government departments. In addition, every department with a significant science budget will be made to appoint its own chief scientific adviser. These advisers will have direct access to their respective Cabinet ministers, and will also help to review one another's research programmes.



Clear winner: Helen Clark does not need Green Party support.

AP

Honey, I shrunk the lab

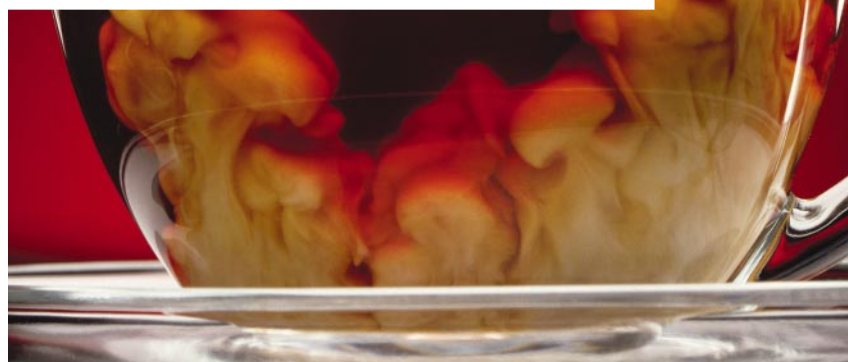
Take a lab full of test tubes, flasks and stirrers, and cram it onto a silicon chip — that's the dream of microfluidics researchers. And as Jonathan Knight finds out, this vision is becoming a practical reality.

About a decade ago, a handful of researchers began discussing an intriguing idea. Could the equipment needed for everyday chemistry and biology procedures be shrunk to fit on a chip the size of a postage stamp? Miniature devices for, say, analysing proteins should be faster and cheaper than conventional versions. Government funding agencies, the military and venture capitalists signed up to the vision, and millions of dollars poured into 'lab-on-a-chip' projects.

But making miniature labs is not just a question of scaling down conventional equipment. Nanolitre volumes of liquid behave in odd ways. For example, we are used to seeing liquids mix by turbulent flow — as illustrated by the way cream swirls into coffee (see right). But such turbulence does not occur in channels just a few tens of micrometres wide. Streams of coffee and cream flow alongside each other over short distances without mixing, almost as if they were separated by glass.

Microfluidics, as the field that deals with this challenge is called, has in recent years made impressive progress in tackling problems such as mixing. Some of the stumbling blocks have been overcome, and basic chips are now on the market. Users confirm that routine tasks can be carried out faster than previously possible. Other problems with miniaturization remain the focus of an intense research effort, which now has its own journal — *Lab on a Chip*, published by Britain's Royal Society of Chemistry.

By definition, microfluidic systems save on reagents. But speed is a more important advantage. The sorting of protein or DNA samples into different sizes by gel electrophoresis, in which an electric field is used to pull the molecules through a gel, takes at least half an hour on the bench but no more than 30 seconds on a chip, for example. And when completely filled with fluid, microfluidic chambers and channels measure volumes more consistently than human hands can, thus reducing error rates.



Fitting a lab onto a chip (top and left) often relies on getting tiny volumes of liquids to mix the way coffee and cream do.

Building the chips is straightforward. The same etching techniques used to make semiconductor chips can create channels and chambers in silicon or glass wafers, which can be layered on top of one other to make more complex patterns. Chip designs can also be stamped or cut into plastic sheets. With the help of computer-design software, researchers can go from concept to prototype in days or hours.

A tight squeeze

But having made a chip, convincing liquid samples to move around the tiny channels is not as easy as it would seem. Air pressure can push samples through channels, but if the sample consists of a concentrated dollop of protein, say, this becomes a problem.

The channel walls exert a drag on the liquid, so that fluid at the centre of the channel moves faster than that at the edge and concentrated samples quickly become smeared.

The most commonly used alternative makes use of a phenomenon called electro-osmosis. In the case of microfluidics, this effect occurs because the channel wall ionizes water molecules near it by donating protons to them. If an electric field is applied along the channel, these positively charged ions flow towards the negative pole and drag the rest of the fluid along with them. Because the pull comes from the edge, the fluid moves as one. Electro-osmosis is widely used in experimental microfluidic chips, and might even turn up in home computers, as Stanford University engineers Daniel Laser and Juan



CALIPER

M. A. BURNS & D. BURKE

D. BISHOP/FOODPIX

Santiago are developing a liquid-cooling system based on electro-osmosis to transfer heat away from microprocessors.

The mechanism behind electrophoresis — the pull exerted by an electric field on charged proteins or DNA — can also be used to move samples around chips. Caliper Technologies, a microfluidics company based in Mountain View, California, makes a gel-electrophoresis chip. Samples are pulled through channels containing a liquid polymer, which sifts biological molecules such as proteins just as conventional gels do. The polymer inhibits electro-osmosis — possibly by clinging to the channel walls — but the pull of the electric field on the sample is enough to load the chip, saturating a 100-micrometre-wide channel with DNA or protein.

Split screen

Once full, the direction of the electric field is shifted so that some of sample is pulled into a side channel where separation can occur. Bands of the DNA or protein, which were stained beforehand with a fluorescent dye, are identified by illuminating the channel with a laser. A single sample can be processed in about 20 seconds.

Persuading liquids in the chips to mix remains one of the biggest problems facing microfluidics researchers. Liquids mix by diffusing into each other. For large volumes, this process can be sped up by stirring, because the turbulence caused increases the area of the interface between the liquids.

But turbulent flow faces opposition in the shape of the viscosity of the two liquids, which tends to keep fluid motion stable. If the sample is sufficiently small, it will never generate the momentum needed to overcome this obstacle. At the microfluidic level, mixing liquids is like trying to stir molasses into honey, and two liquids travelling side-by-side through a narrow channel only become fully mixed after many centimetres.

One approach to accelerating the process is to stretch and fold the fluid layers as they move down the channel, like kneading a ball of dough. In January, George Whitesides' group at Harvard University achieved this using a herringbone pattern of ridges on the channel's floor¹. The ridges rotate the flow clockwise on the left side and anticlockwise on the right, causing fluid at the centre to fold in on itself. The mix is complete in only 3 cm — short enough to fit on a chip if a winding

channel is used. Without the ridges, the liquids would only have mixed after 100 cm.

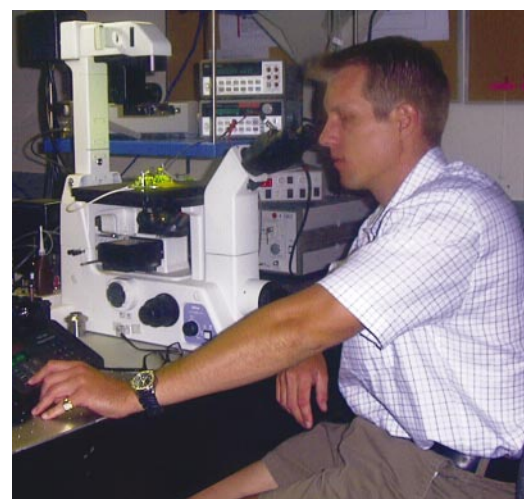
Michael Oddy, a microfluidics engineer in Santiago's Stanford lab, discovered another mixing technique while experimenting with electro-osmosis. He found that applying an alternating current causes the fluid to oscillate in the channel, and that the oscillation is unstable at certain frequencies². The details of why this happens remain unclear, but mixing times are much reduced. "We get a nearly homogeneous mixture in about 3 seconds, compared with 500 seconds for diffusion alone," says Santiago.

Other researchers are finding that the lack of mixing can sometimes be an advantage. Paul Yager, a bioengineer at the University of Washington in Seattle, has developed a device for measuring the concentration of solutions of small molecules such as peptides³. In his chip, two solutions flow into the two arms at the top of a T-shaped channel 1,200 micrometres wide. One solution contains the molecule of interest and the other a larger antibody that binds to it.

Separate streams

When the liquids meet and head down the main channel, the lack of turbulence ensures that mixing occurs only by diffusion. But the few smaller molecules that diffuse into the neighbouring layer bind to the much larger antibody molecules, which effectively halt the movements of the molecules they capture. The higher the concentration of small molecules, the more of them escape capture and are able to diffuse deeper into the antibody solution. By tagging a few of the small molecules with a fluorescent label, the diffusion rate, and hence the concentration, can be revealed by illuminating the antibody solution with a laser.

But making a true lab on a chip means combining several processes. At the University of Michigan, chemical engineer Mark Burns has produced an electrophoresis chip that does just that⁴. Two parallel channels hold precise volumes of liquid, such as a stained solution of DNA and an enzyme that can cleave DNA. Air pressure is used to push the samples into a chamber, where a micro-electronic heater warms them to the optimal reaction temperature. They are incubated for about 15 minutes, and then loaded into a channel containing a gel for electrophoresis.



A whole lotta shakin': Michael Oddy uses an alternating current to mix liquids by oscillation.

Finally, a photodetector notes fluorescent bands of DNA fragments in the gel corresponding to specific DNA sequences that were present in the original sample.

For some users, there is no need to cram all of this hardware onto the chip. The Caliper chips, for instance, use a toaster-sized instrument to house the laser and the electronics to run the electrophoresis. Pharmaceutical researchers, for example, don't care if the equipment fills the room, as long as it is fast and accurate, says Andreas Manz, a chemist at Imperial College London, and one of the pioneers of microfluidics.

But others see portability as an essential part of the future of microfluidics. Yager, whose concentration sensor is being commercialized by Micronics, a Redmond-based company that was spun off from the University of Washington in 1996, sees a huge market in home healthcare. He suggests that his device could be used to conduct blood tests with only a pinprick sample, for example.

Burns, meanwhile, believes that his chips could ultimately be developed into a home kit to test for bacterial DNA. Users could plug the chips into a laptop, and e-mail the results to their doctor. "We're building the expertise into the chip," says Burns. If such predictions turn out to be correct, the new chips won't just be shrunken labs — they will also embody the skills of laboratory technicians. ■

Jonathan Knight writes for Nature from San Francisco.

1. Stroock, A. D. *et al. Science* **295**, 647–651 (2002).
2. Oddy, M. H., Santiago, J. G. & Mikkelsen, J. C. *Anal. Chem.* **73**, 5822–5832 (2001).
3. Hatch, A. *et al. Nature Biotechnol.* **19**, 461–465 (2001).
4. Burns, M. A. *et al. Science* **282**, 484–487 (1998).

Lab on a Chip

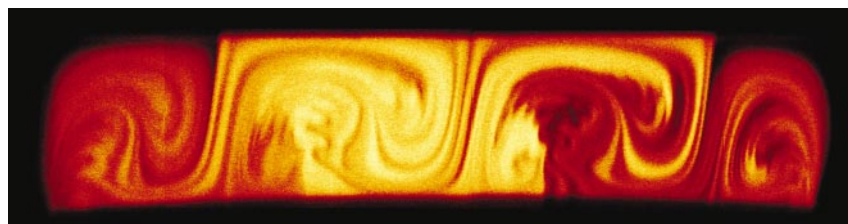
► www.rsc.org/isis/journals/current/loc/locpub.htm

Caliper Technologies

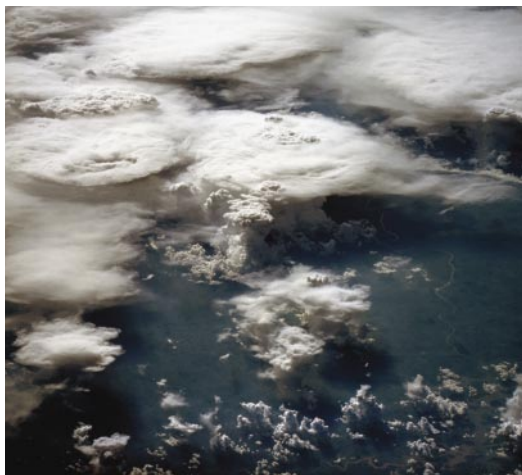
► www.calipertech.com

Micronics

► www.micronics.net



Meeting a knead: small ridges in microchannels can force liquids to mix by folding them together.



NASA

When doubt is a sure thing

Is it possible to adopt a more rigorous approach to the communication of scientific uncertainty? Jim Giles talks to the climatologists whose pursuit of this goal has seen them dubbed the 'uncertainty cops'.

Stephen Schneider is very certain in his views about uncertainty. As a leading climate researcher at Stanford University in California, he has plenty of experience in dealing with it. Global temperatures are set to rise — but no one is sure by how much. And although this warming will influence variables from biodiversity to economic productivity, exactly how it will do so remains unclear.

Schneider cannot eliminate uncertainties from his work, but he has little time for fellow scientists who are sloppy in expressing just how unsure they are. He argues, for instance, that the phrase "low confidence" should have a precise, quantitative meaning. He espouses the use of graphical tools to illustrate where scientific uncertainties come from. And he is adamant that statistical confidence levels should be attached to even the most complex scientific predictions.

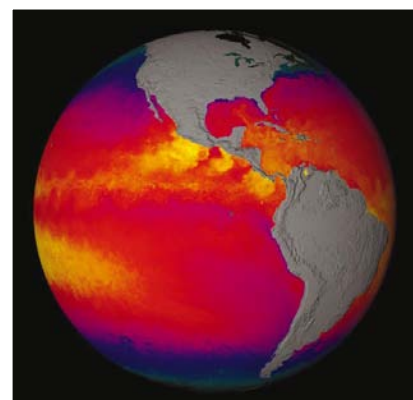
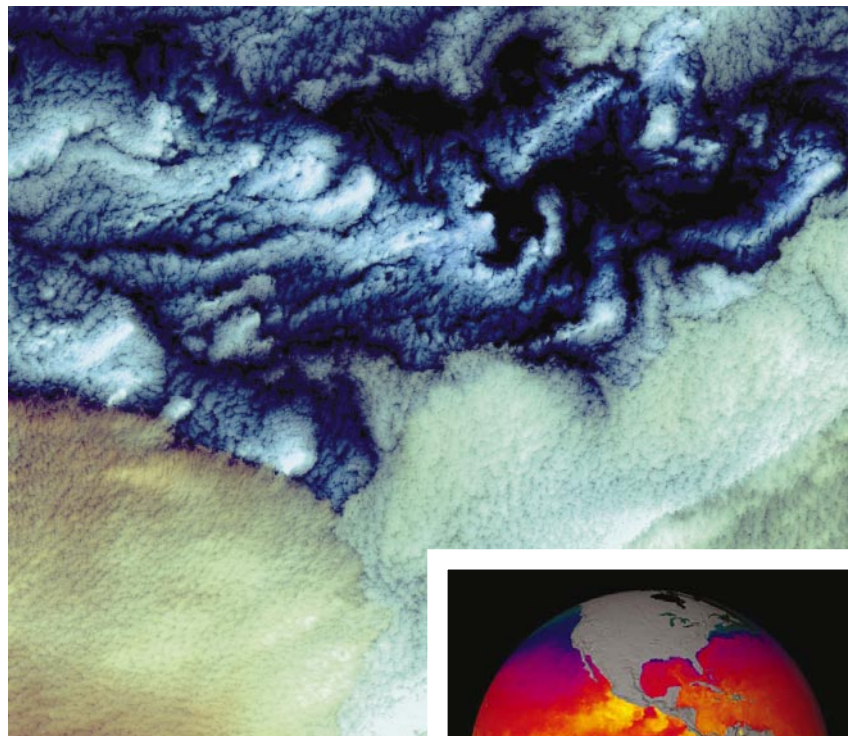
Together with Richard Moss, currently executive director of the Global Change Research Program in Washington DC, which coordinates the US government's climate-change research, Schneider has been pestering his colleagues about these ideas for more than five years. With plans for a major new international climate assessment just getting started, Schneider and Moss are gearing up to pro-

mote their agenda once again. They also believe that other subjects that engender both public concern and scientific uncertainty could benefit from taking a similar approach.

Schneider and Moss began their campaign after witnessing the furore that greeted the 1995 publication of the second climate assessment by the Intergovernmental Panel on Climate Change (IPCC). The report was the first IPCC publication to state unequivocally that human activities are having a discernible impact on the Earth's climate. Environmental pressure groups seized upon this statement, while lobbyists for the fossil-fuel industry and a minority of sceptical climate scientists complained bitterly that uncertainties in the science had been played down.

Dazed and confused

In part, claim Schneider and Moss, the rancour stemmed from the confused way in which the report dealt with uncertainties. Moss cites the example of its estimate of climate sensitivity — the increase in average global temperatures that is expected to occur if carbon dioxide concentrations were to reach twice their pre-industrial-revolution levels. This was given as a range of 1.5–4.5 °C. Environmental groups naturally quoted figures at the top of the range, which are



Clouded world view: data on future climate trends are shrouded in scientific uncertainty.

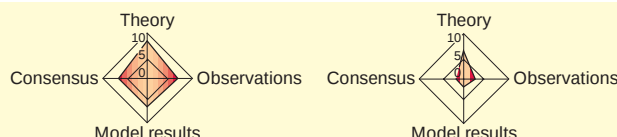
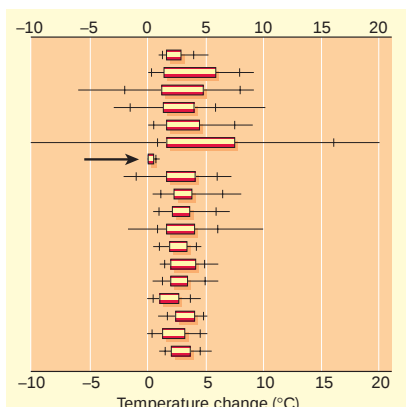
actually less likely to materialize than those near the middle. This point, which was often missed in the arguments that followed the report's publication, would have been much clearer if probabilities had been assigned to different parts of the range, says Moss.

Reluctance to use probabilities in this way is common in climate-change research, and stems from the nature of the predictions involved. In related fields, results that can be calculated in a straightforward manner often have probabilities attached to them. Weather forecasters, for example, use computer models to predict whether certain atmospheric conditions will bring rain or shine. Because their models have been tested using real data from past weather systems, the forecasters know how accurate their predictions are.

But models that simulate long-term climate changes have no comparable reality check, so climate modellers often shy away from using probabilities. Instead, they qualify their results by discussing how well understood a model is, or by noting the availability of observational data. It was precisely this qualitative approach, argue Schneider and Moss, that led to the arguments over how the

USGS

J. ALLEN/MODIS/UNIV. MIAMI/RSMs



Cops on patrol: Richard Moss (top, left) and Stephen Schneider want climatologists to use four-axis plots to show where the uncertainty in their predictions comes from. The examples (above) are for temperature change (left) and flooding risk. The pair also advocate plotting individual scientists' predictions (top right) — this example shows that one expert (arrowed) used different methods.

1995 IPCC report should be summarized.

Schneider and Moss tried to make sure that this mistake was not repeated in the IPCC's next report. In 1996, they held a session on the presentation of uncertainty at the Elements of Change conference at the Aspen Global Change Institute in Colorado. They also consulted risk-communication experts and pressured the IPCC's leaders to recognize the problem. The result was a paper¹, published by the two 'uncertainty cops' in 2000 as the IPCC prepared its huge third assessment for publication the following year.

Confidence boost

At the heart of Schneider and Moss's paper lay a call for researchers to use a common language when describing their results. The pair argued that the phrase "low confidence", for example, should only be used in connection with confidence ratings of between 5% and 33%. In all, Schneider and Moss recommended five categories, ranging from "very low confidence" (less than 5%) to "very high confidence" (95–100% confidence).

If a lack of data prevents uncertainty from being calculated using standard statistical techniques, Schneider and Moss suggested that authors should assign subjective probabilities to their results, taking into account their knowledge of the models and data behind them. This method, known as bayesian statistics, is not as woolly as it sounds. Initial confidence levels may be subjective, but they can be modified as more data about a system become available. The degree to which the initial assumption biases the final confidence level can also be assessed. "Bayesian

statistics allow scientists to indicate their degree of belief in a result given the information available to them," explains Moss.

For those who are uncomfortable with using subjective estimates, Schneider and Moss suggested a descriptive scale for assessing the "state of knowledge" about a result. The term "well established", for example, should describe a result for which the models involve known processes, the observations are largely consistent with the models, or the finding is supported by several lines of evidence.

Did the IPCC's authors listen? Well, yes and no. When the third assessment was published last year, the panel's Working Group I, which focuses on the atmospheric and oceanographic science behind climate change, embraced the terminology for describing confidence levels, and even added a new category — greater than 99% confidence was termed "virtually certain". Working Group II, which seeks to assess the impact of climate change on ecosystems and human activities, used the scale in its original form — perhaps unsurprisingly, as Schneider was one of the lead authors of the group's report. But Working Group III, which is charged with outlining techniques for tackling climate change, did not use the terminology at all.

Schneider ascribes this experience to differences in culture between the disciplines that make up the working groups. Many of the climate researchers of Working Group I were used to dealing with predictions that carried high degrees of confidence, and so were comfortable with the ratings — and were happy to

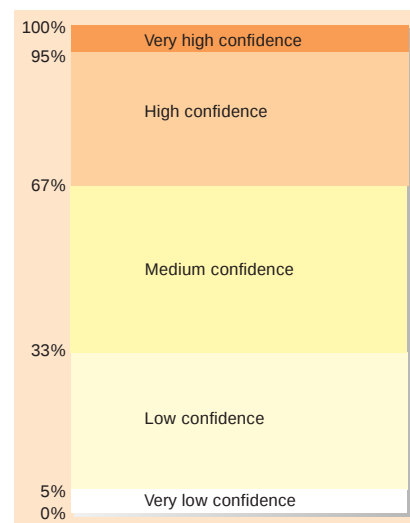
add another at the top of the scale. But such high confidence is rarely shared by the sociologists and economists who were involved in the other two working groups, hence their more circumspect approach.

Many of the IPCC's authors — particularly the members of Working Group I — were suspicious of bayesian statistics, preferring instead to rely on Schneider and Moss's qualitative scale when data from similar past events were not available. These attitudes also stem from the authors' backgrounds, argues Schneider. "The economists understood it was not irresponsible," he says. "But it was tougher for your average natural scientist." Mike Hulme, executive director of the Tyndall Centre for Climate Change Research at the University of East Anglia in Norwich, UK, agrees. "It is not a normal approach in the lab," he says. "But it is useful when communicating results to policy-makers."

Get in shape

Schneider and Moss have also developed graphical techniques to summarize how confidence, or lack of it, arises. One scheme involves plotting points on four axes — corresponding to confidence in the theory, the observations, the models and the consensus within a field — arranged like the points on a compass. The points are then joined to create a shape — its area indicates the overall degree of confidence in the result, and its outline describes how that confidence arises (see figure, above left). The technique was used by Mike Scott of the Pacific Northwest National Laboratory in Richland, Washington, in his section of Working Group II's report dealing with the impact of climate change on human settlements. "I could kiss him," says a clearly delighted Schneider.

Other graphical techniques in Schneider and Moss's paper were intended for use in the absence of clear consensus over a result — such as for climate-sensitivity estimates.



Schneider and Moss's proposed confidence bands.

Rather than averaging these estimates, Schneider and Moss proposed a graphical approach that was used in a 1995 paper by climate-policy experts Granger Morgan, of Carnegie Mellon University in Pittsburgh, and David Keith, then at Harvard University².

Morgan and Keith asked several climate experts to estimate the range of climate sensitivity, and found that one of them made a very different prediction from those of the others. Averaging these results would have obscured the fact that this researcher was using different assumptions and models to generate the prediction (see figure, previous page).

Morgan–Keith plots were not used in the IPCC's 2001 report. But Schneider and Moss say that they will be pushing for their ideas to be used more consistently when researchers begin work on the panel's fourth assessment, due to be published in 2007. Moss is particularly keen for researchers to adopt another of his proposals — he wants complex predictions depending on a series of results to come with a “traceable account” that includes details of what lines of evidence were used to generate each result, and the uncertainties associated with that evidence.

Estimates vary

But the uncertainty cops will have to overcome some resistance — as Schneider found last time around. Before producing its third report, the IPCC asked a group of academic researchers, environmentalists, industry representatives, engineers and economists to consider how emissions of greenhouse gases are likely to change over this century. These emissions scenarios were used in parts of the third assessment — Working Group I used them to produce estimates of how different emission levels would affect the climate, and Working Group II used the resulting estimates to consider the likely impact of the predicted climate changes.

At a meeting of the IPCC's scenarios task group, held in 1997 at the International Institute for Applied Systems Analysis in Laxenburg, Austria, Schneider attempted to persuade delegates to assign relative probabilities to the emission scenarios they had conceived. The idea was controversial, as emissions depend on many factors, from new technologies to changes in population. Delegates held widely different views on these issues, and most believed that simply identifying the various scenarios was as much as they could do.

That's all very well, says Schneider, but it means that policy-makers can assume that all emissions scenarios are equally likely. Special-interest groups can then use whichever scenario fits their agenda when making claims about climate change, and politicians can pick those that help to justify their policies. So Schneider says that he will continue to push for probabilities to be attached to the IPCC's emissions scenarios.

Given that climate change is just one of



On shaky ground: researchers studying GM crops and seismic data are also plagued by uncertainty.

as the possibility that introducing genes for virus resistance could lead to the development of super-weeds,” he argues. But he says that it is difficult to think of a single variable that can be used to represent these fears, as would be required for a Morgan–Keith plot. “The method may be useful for precisely phrased issues,” says Perry, “but GM debates have rarely restricted themselves in this way.”

Perry is also concerned about Schneider and Moss's four-axis plots. He warns that the area of the plotted shape may not accurately represent the overall confidence in the result if the uncertainties associated with the different axes are not independent. “Theory comes from observations, and both of these feed into models, so the different axes may depend on each other,” says Perry. “If they are dependent, the size of the shape will not be representative of the total probability.”

Moss admits that the theory behind the four-axis plots is not as precise as that which underlies more established areas of statistics. He says that he and Schneider are “blazing a bit of a trail”, adding that “this kind of process is new”. Rather than seeing the plots as the finished article, Moss hopes that they will evolve after being used by other researchers.

Schneider accepts that he has some way to go in convincing his peers to embrace his approach to the communication of scientific uncertainty. But he certainly cannot be accused of failing to practise what he preaches. When he was diagnosed with lymphoma, Schneider says that he used the Bayesian approach to help make decisions about his various treatment options. “My recovery is doing great,” he adds.

With many senior politicians still unconvinced by the IPCC's consensus on global warming, Schneider and Moss believe that it is more important than ever to get the world's climate scientists signed up to their uncertainty manifesto. “There is no other way to advise policy-makers,” claims Schneider. ■

Jim Giles is Nature's associate news and features editor.

1. Moss, R. H. & Schneider, S. H. in *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC* (eds Pachauri, R., Taniguchi, T. & Tanaka, K.) 33–51 (World Meteorol. Org., Geneva, 2000).
2. Morgan, M. G. & Keith, D. *Env. Sci. Tech.* **29**, 468–476 (1995).

DNA points the way ahead in taxonomy

In assessing new approaches, it's time for DNA's unique contribution to take a central role.

Sir — The overwhelming task for taxonomy in ecological and biodiversity research arguably requires entirely new approaches. Web-based technology would be a great step towards a more accessible and universal platform, as proposed by H. C. J. Godfray in his Commentary¹, and backed by F. A. Bisby in Correspondence². But some crucial problems remain, not least the quality and accuracy of submitted information³. Many of these problems could be solved if DNA sequences were used as the universal reference standard.

Their usefulness for taxonomic purposes is not disputed. However, all current approaches use DNA as just an additional criterion for identifying a species or a taxon, without attempting to give it a central role. We believe that DNA taxonomy can provide a new scaffold for our accumulated taxonomic knowledge and a reliable tool for species identification and description.

DNA sequences alone are not sufficient to characterize a species⁴, but their unique reproducibility helps to guard against duplicate descriptions³. Moreover, collection and curation of extracted DNA

samples is technically easy. DNA is very stable and any sample can be split into multiple subsamples, which can be sent to other museums as backups. DNA collections are already needed for the many projects in different laboratories looking at the phylogeny or phylogeography of species. These projects often involve very valuable samples, yet there is no scheme to safeguard them for future reference.

DNA sequencing is often considered to be a complex, expensive technology. But training good taxonomists is also very expensive, and it is a waste of resources to use them only for routine identification of specimens collected in research projects. We need good taxonomists to work on the huge task of matching existing taxonomic information with DNA sequences of new specimens, and to recognize and describe new species. Routine species identification should be the task of specialized DNA sequencing facilities, for which machines are readily available. A DNA facility that could routinely handle about 1,000 samples a day would cost approximately as much as a facility running a transmission and a scanning electron microscope.

We believe that a system based on DNA taxonomy can now be built to integrate the strengths of the traditional system with new technological possibilities, making full use of the invaluable information accumulated over the centuries and giving well-trained taxonomy specialists the opportunity to convert their expertise into broadly reproducible knowledge. We are developing these ideas into a longer article to be published elsewhere soon.

Diethard Tautz*, Peter Arcander†, Alessandro Minelli‡, Richard H. Thomas§, Alfred P. Vogler||

*Institut für Genetik der Universität zu Köln, Weyertal 121, 50931 Köln, Germany

†Department of Evolutionary Biology, Institute of Zoology, University of Copenhagen, Universitetsparken 15, DK-2100 Ø Copenhagen, Denmark

‡Department of Biology, University of Padova, Via Ugo Bassi 58 B, 35131 Padova, Italy

§Department of Zoology, The Natural History Museum, Cromwell Road, London, SW7 5BD, UK

||Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

1. Godfray, H. C. J. *Nature* **417**, 17–19 (2002).

2. Bisby, F. A. *Nature* **418**, 367 (2002).

3. Lee, M. S. Y. *Nature* **417**, 787–788 (2002).

4. Ferguson, J. W. H. *Biol. J. Linn. Soc.* **74**, 509–516 (2002).

Israeli concern about Palestinian suffering

Sir — I am an Israeli chemist and a human-rights volunteer writing in response to your Opinion article “Collaboration in extremism” (*Nature* **417**, 207; 2002), in which you expressed encouragement and support for joint Israeli–Palestinian collaborative science. (See also Correspondence, *Nature* **417**, 689–690; 2002.)

Although well-intentioned, your article is, in my view, off the mark. The situation of the 3.5 million people in Palestine is much worse than you imagine. People are locked in disconnected ghettos, sometimes invaded by the Israeli army. Without a state or citizenship of any country, the Palestinians are unprotected by a legal system, and totally in the power of an army run by people whose notion of ‘national security’ is to wreak as much destruction in Palestine as they can get away with.

Travel between ghettos is difficult and dangerous, sometimes deadly. External help is hard to get; journalists and human-rights activists are routinely barred by the simple expedient of declaring ‘closed military zones’. There is no end in sight, as most Israelis seem unable to grasp that Sharon’s policies promote bloodshed on both sides instead of preventing it.

So the laudable suggestions of your article, to promote normalization and collaborations, do not seem relevant. At present, Palestinian academics are unlikely to place high priority on the quality of their scientific collaborations. They are more likely to worry about issues such as being able to get to work tomorrow, being sent to a detention camp, the next military invasion, demolition of their homes, harassment at military roadblocks and, indeed, about survival.

Friends in science, help by well-meaning foreigners is badly needed here. However, please keep it to the point.

Victoria Buch

Israel (full address supplied)

Science council replies to neutron claims

Sir — Last week in Correspondence (*Nature* **418**, 367; 2002) you published the views of seven subcommittee members of Germany’s science council, the Wissenschaftsrat, about the European Spallation Source (ESS).

The council released a statement on nine large-scale facility projects in July this year. These include the ESS, which was evaluated by the subcommittee in a

meeting last December. The draft version of their final statement prepared by the Wissenschaftsrat’s secretariat was based on the conclusions agreed at that meeting. In April, all subcommittee members were given three weeks to comment on this draft version. With some modifications and additional comments, all except one endorsed the statement on the relevance of the scientific case for the ESS, which some of them are now disputing.

The Wissenschaftsrat considered the overall importance of each of the nine projects in terms of science policy on the basis of each subcommittee’s findings. It goes without saying that a science-policy assessment can come to different conclusions from those of a disciplinary evaluation. Funding decisions will be based on the Wissenschaftsrat’s comprehensive statement which takes several subcommittees’ reports into account.

As chair of the Wissenschaftsrat, I was surprised to read that a statement recently agreed on by all members of a subcommittee except one is now drawn into question by several members. I would find it more helpful if the debate focused on the scientific merit of the ESS project.

Karl Max Einhäupl

Wissenschaftsrat, Brohlerstrasse 11, 50968 Cologne, Germany

A whole new world of geology

The canals have faded from view, but Mars remains tantalizingly out of reach.

Mapping Mars: Science, Imagination, and the Birth of a World

by Oliver Morton

Fourth Estate: 2002. 351 pp. £18.99

Greg Bear

Imagination is the one thing in our exploration of Mars that has never failed us. We imagine ourselves on Mars. We imagine our descendants there. Some of us even stake our careers and livelihoods on a particular interpretation of Mars, just as the US astronomer Percival Lowell did in the years surrounding the birth of the twentieth century.

I am very short-sighted. When I look without my glasses over a mottled view of tree branches, I see scalloped and arcuate shapes, doubled lines, circular blobs of light intersecting and competing for my unfocused attention. That the martian surface even today shows arcuate and scalloped features is well known — but no matter how hard I squint, half-blind, I cannot make a map or photograph of Mars show lineaments, either single or doubled. What did Giovanni Schiaparelli, Lowell, Andrew Douglass and the other nineteenth-century astronomers see on those long cold nights of viewing? Tree branches? Very unlikely, although I can certainly duplicate doubled canals as I peer myopically at long, bare twigs. Eyelashes, perhaps? Floaters, those chains of cells crossing the visual field? Or real lineaments that have since mysteriously vanished?

Whatever they saw, Lowell's theories of Mars as a marginally inhabitable, dying planet drove the human race Mars-mad. They gave the red planet a compelling personality, a history, an architectural style and an air of neurasthenic decadence.

Oliver Morton's excellent study of Mars and the human imagination — scientific, psychological and literary — gives a broad overview of martian discovery and description. He focuses on geology and mapping, complete with personalities, places and hypotheses both wrong-headed and brilliant.

Mapping Mars admits us into the bars and bedrooms of science, where scientists flirt with science fiction, and science fiction writers hang around the Von Karman Auditorium at the Jet Propulsion Laboratory (JPL) in Pasadena, California, like jealous suitors competing for favours. I missed the Viking voyages, but I was there for the long run of Voyager missions, along with many other writers, artists, politicians and even movie stars. We moved to the Pasadena Civic Center for the triumphant Mars Pathfinder and the disappointment of the Mars Polar



A romantic view of Mars: Chesley Bonestell's paintings capture the magic and mystery of the red planet.

Lander. Science came alive, and scientists and laymen rubbed shoulders to the benefit of both. We all became amateur geologists. The social interactions that Morton depicts ring true, and the personalities I remember so vividly from press conferences and conversations are deftly rendered.

Through so much of our romance with Mars, we've been haunted by canals. The lineaments or canals of Mars are explicit in Chesley Bonestell's paintings of the 1940s and 1950s. Even in 1965, a NASA/JPL schematic of Mars, putting in perspective the 22 squares photographed by Mariner 4, showed long canals.

Clearer pictures of Mars from Mariner and Viking changed all that. No one has captured canals on photographs and they no longer appear on maps of Mars. Most scientists believe that they never existed — but Mars may yet prove us wrong. Our dreams and theories draw closer to the truth year by year, but still without that tantalizing and essential element: actually walking on Mars.

Geologists, Morton tells us, are frustrated by looking at rocks in pictures. They long to look at the same rock from many angles, strike it with a pick, lift it and rub it against other rocks. Without these hands-on clues, highly disciplined speculation remains the best arbiter of the disputes provoked by better and more detailed evidence.

The geological histories of Earth and Mars ran in parallel. The scablands in Oregon provide clues to similar features on Mars. Searching for solutions to what appear at first to be uniquely martian problems, scientists turn to the neglected geographic features of Earth, and come up with hypotheses that may explain both.

From Bruce Murray's enthusiastic embrace of a dead, dry Mars in the 1970s, we now embark on a conceptual period in which Mars almost sloshes. The red planet seems awash with secret water, with or without carbon dioxide fizz. Meteor strikes, earthquakes or volcanism could have unleashed devastating torrents from ice-capped aquifers;

ice dams on rapidly forming rivers or lakes behind dust-hidden glaciers could have burst forth; water could reappear as brief dews or as a build-up of thousands of centuries of light snow. Clathrates may release huge volumes of atmospheric gas, altering conditions for centuries.

The idea of a seasonal Mars comes and goes, itself weirdly seasonal. What is impossible one year suddenly becomes likely the next. The ferment of theorizing leads to private and public confusion, but Morton pulls us along with skill and insight through bull session after late-night bull session and conference after conference, keeping the players straight, engaging in light gossip, but clearly explaining the real issues and their ramifications.

The result is one of the best popular-geology books I've read, a match for John McPhee's *Basin and Range* (Farrar Straus Giroux, 1982). To geologists, *Mapping Mars* will probably function as both a solid lunch and a plate of dessert nibbles, both food and sweets for thought.

Closing the book, science passes into day-dream, made vivid by well-drawn scenarios. Mars may indeed be a heaven for geologists. It is the most nakedly fascinating and possibly the most geologically complicated world in our Solar System. Soon, a geologist will go to Mars and pick up a rock, strike it against another rock, and see up close what only human eyes can see. Something surprising, probably impossible.

The fun has just begun. ■

Greg Bear has both painted and written about Mars. He is the author of *Moving Mars*, *Darwin's Radio*, *Vitals* and the forthcoming *Darwin's Children*, and lives in Lynnwood, Washington, USA.

A historian of Victorian astronomy

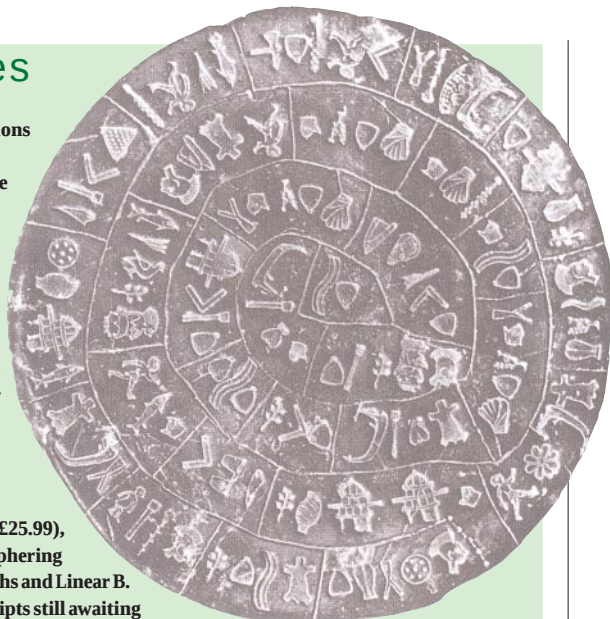
Agnes Mary Clerke and the Rise of Astrophysics
by Mary Brück
Cambridge University Press: 2002. 286 pp.
£35, \$50

Donald E. Osterbrock

Agnes Clerke was the most influential writer on astronomy in the English language at the end of the Victorian era. In those days before the Internet, television, radio and even newsreels, people kept abreast of the latest developments in the world by reading books, magazines and newspapers. Astronomy — the study of the whole Universe — was as interesting then as it is now. Numerous popularizers churned out repetitious pot-boilers with such titles as *In Starry Realms* and *Story of the Heavens*, but Agnes Clerke wrote solid, well-researched, up-to-date reviews and books for the educated public.

Lost languages

Writing is among the great inventions in human history — perhaps the greatest invention, because it made history possible. The process of deciphering ancient scripts is an intellectual and imaginative challenge to both archeological scholars and amateurs alike. Each script is unique, so decipherment is akin to invention. These ancient writing systems provide a window to past civilizations, and help us to understand how our modern writing systems function. Andrew Robinson, in his new book *Lost Languages* (McGraw-Hill, \$34.95, £25.99), tells the fascinating stories of deciphering Egyptian hieroglyphs, Mayan glyphs and Linear B. He then surveys the important scripts still awaiting decipherment, including the Indus script and the Phaistos disc of Crete (above), believed to be the world's oldest 'printed' document, dating back to before 1600 BC.



Her three most important books, *A Popular History of Astronomy in the Nineteenth Century* (1885), *The System of the Stars* (1890) and *Problems in Astrophysics* (1903), were authoritative surveys of the subject and its history. Probably even more important were the hundreds of articles she wrote, many of them for *The Edinburgh Review*, a quarterly review actually published in London throughout her lifetime. From 1877 to 1907, when her last article was printed after her death, Clerke had kept her readers informed and interested in the progress of astronomy and astrophysics.

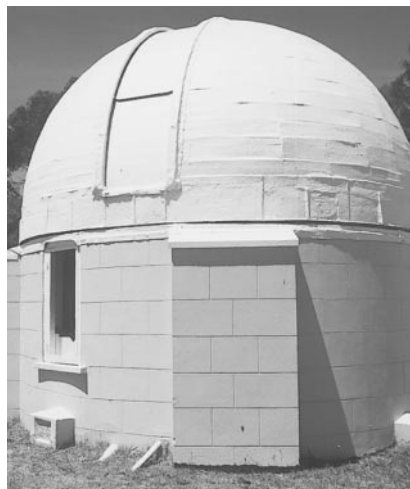
Mary Brück's book tells the story of this remarkable woman. Agnes Clerke never attended school, nor studied at a university. Born in the small town of Skibbereen in County Cork, Ireland, she was the daughter of a mixed marriage between a Protestant bank manager and a daughter of a wealthy Roman Catholic family. Agnes was educated at home until the family moved to Dublin in 1861, when she was nineteen. There she continued her studies in the city's libraries. Six years later the family moved to Italy, where they stayed for ten years. Agnes was an excellent linguist, and by the time they all moved to England she was fluent in Italian, French, German and Spanish, as well as Latin and Greek. And she was very well read, especially in science.

Clerke had already written two articles for *The Edinburgh Review* from Florence, one describing the rise of the Mafia, and the other on Copernicus in Italy. After moving to London, she soon established herself as a dependable writer with a lively style who could tackle almost any subject in science or classics and produce a well-written article on schedule. She became a contributor to the *Encyclopaedia Britannica*, which was

then being revised, beginning with a long article on Galileo.

Friends were soon urging Clerke to write a history of astronomy, and she realized that she could do it. Astronomy was changing fast at the time, as astrophysics came to the fore. She studied physics and astrophysics on her own, and in four years produced her first book, *A Popular History of Astronomy in the Nineteenth Century*, which established her as an expert, not only with general readers, but also with astronomers, several of whom reviewed the book very favourably.

She began consulting the researchers themselves to learn their latest results. She had already written to Edward Holden, later the first director of Lick Observatory and himself a prolific writer. He had encouraged her; now all the astronomers she interviewed



Observation point: Agnes Clerke made her own studies of the stars on a visit to South Africa.

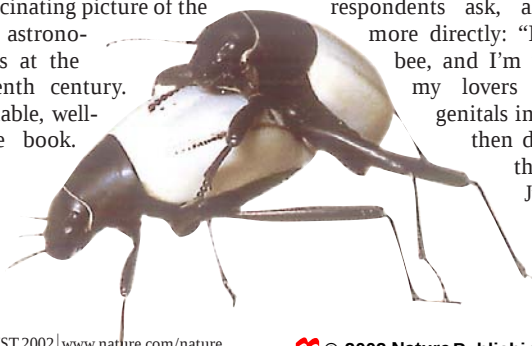
or wrote to were glad to help, particularly with information about their own discoveries and triumphs. David Gill, a Scottish astronomer who observed the southern skies from South Africa, became a particularly important source for her. In London she met and talked frequently with William Huggins, a pioneer English astrophysicist, and with Norman Lockyer, the founder of *Nature*. At first she featured their work, but later she came to believe that Huggins was past his prime and that Lockyer claimed more than he actually proved. She paid for that by getting negative reviews of her later books in *Nature*. Pushy Americans such as W. W. Campbell and young George Ellery Hale flooded her with their results, and she found some of them important enough to be included in her books.

Clerke also kept up-to-date by attending scientific lectures in London. As a woman she was not permitted to join the Royal Astronomical Society, nor to attend its monthly meetings, but she entered its precincts to use its library and to view exhibitions of recent results, especially astronomical photographs. She wrote many articles for the *Observatory*, the astronomical magazine closely associated with the society, although not owned by it. She was later allowed to come to the society's meetings as a guest of a member, and towards the end of her life she was elected as an honorary member.

Brück has written an excellent book. Her thorough research in numerous archives and in Clerke's publications is well documented in the notes, and numerous well-chosen photographs illustrate the text. There is an extremely informative chapter on women in astronomy in the Victorian era. She concludes, no doubt correctly, that a few women could get better jobs in astronomy in America than in Britain, but it seems to me that the difference was only marginal. The case of Alice Everett, a Cambridge graduate, illustrates this well. She did good work as an assistant at Greenwich, Potsdam and Vassar College in the United States, but couldn't get a permanent position at any of them. She also applied unsuccessfully to the Lick Observatory, where the director, James E. Keeler, was searching for a male observer at the time.

Brück paints a fascinating picture of the rich fabric of British astronomy and astrophysics at the end of the nineteenth century. This is a highly readable, well-produced, attractive book.

■ Donald E. Osterbrock is at the Lick Observatory, University of California, Santa Cruz, California 95064, USA.



A bestial feast

Dr Tatiana's Sex Advice to All Creation: The Definitive Guide to the Evolutionary Biology of Sex by Olivia Judson
Metropolitan Books/Chatto & Windus: 2002.
320 pp. \$23/£16.99

Tim Birkhead

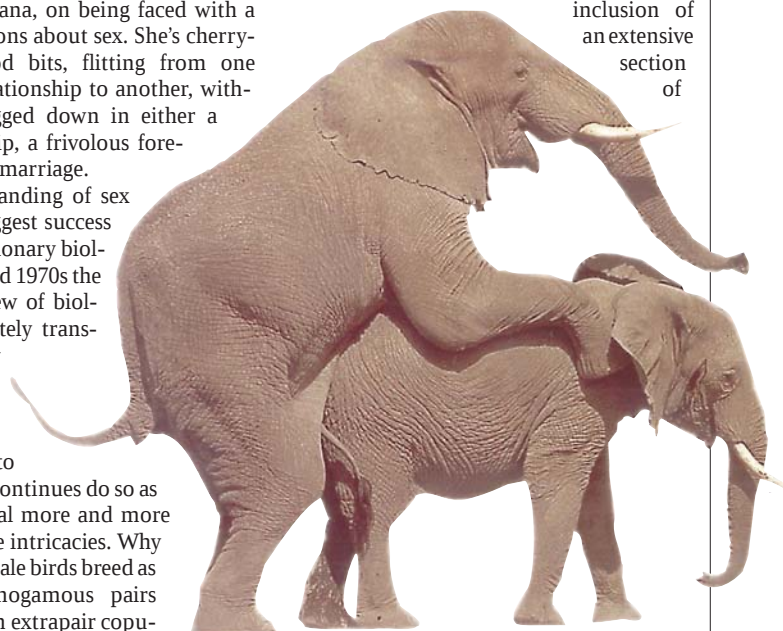
The first time I went to an international conference and stayed in an expensive hotel, I was amazed by the veritable wall of food that confronted me at breakfast the next morning. It was an optimal forager's fantasy and I creamed off all the tasty bits from a wide range of dishes. As far as I could tell, everyone else did the same. And that is precisely what Olivia Judson has done, under the guise of agony-aunt Dr Tatiana, on being faced with a surfeit of questions about sex. She's cherry-picked the good bits, flitting from one extramarital relationship to another, without getting bogged down in either a lengthy courtship, a frivolous foreplay or a tedious marriage.

Our understanding of sex is one of the biggest success stories in evolutionary biology. Since the mid 1970s the 'selfish gene' view of biology has completely transformed our view of all aspects of reproduction, from anatomy and physiology to behaviour, and continues to do so as researchers reveal more and more of its remarkable intricacies. Why do male and female birds breed as apparently monogamous pairs yet still engage in extrapair copulations? Why do some flatworms have a dozen penises when some of us have to make do with one? Why does the female spotted hyena possess a clitoris the size of a courgette, only to have it ripped apart when she gives birth? These are the kind of questions that Dr Tatiana's concerned correspondents ask, albeit rather more directly: "I'm a queen bee, and I'm worried. All my lovers leave their genitals inside me and then drop dead. Is this normal?"

Judging from her replies I imagine Dr Tatiana to be some-

thing like Dame Edna Everage with a PhD. Nonetheless, as every agony aunt should, Dr Tatiana proffers light-hearted, informative advice. Sex is fun, and fun to read about, especially as in this case it is presented in bite-sized pieces that you can snack on as the fancy takes you. What's on offer spans the entire animal kingdom, from rats to rotifers, and bacteria to birds, and somewhat less digestibly includes some theory too.

The danger, of course, is that presenting research findings as a kind of buffet can trivialize the entire process, turning it into a fast-food fix with none of the benefits of a balanced meal. But Judson walks this particular hedonistic tight-rope with aplomb — her text is both wonderfully entertaining and authoritative — and the inclusion of an extensive section of



notes and references ensures that she retains her scientific respectability. All in all this is a stimulating feast of extraordinary sexual practices, and my advice is to place *Dr Tatiana's Sex Advice* in the smallest room, where it will keep you entertained and informed for hours.

■ Tim Birkhead is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.



Completing the circle

P. G. Schultz and R. A. Lerner

Historically, immunochemistry has focused on understanding and exploiting the remarkable binding affinity and specificity of antibody molecules. With the advent of antibody catalysis, chemists could ask for the first time whether this sophisticated system of molecular diversity can be used to create new chemical tools — efficient, selective catalysts. From these experiments have arisen a host of antibody catalysts, in some cases with specificities and activities that rival those of enzymes.

Moreover, characterization of the mechanisms and immunological evolution of catalytic antibodies has provided fundamental insights into the evolution of binding and catalytic functions in nature. The field has also had a broader influence on chemistry — chemists are increasingly incorporating the biological idea of molecular diversity into their efforts to synthesize molecules with new functions. Finally, the study of antibody catalysis has led to the realization that all antibodies are oxidative catalysts, suggesting a previously unrecognized role for the antibody molecule in the immune response.

The classic early theories of enzyme catalysis by Linus Pauling and J. B. S. Haldane focused on proteins' ability to use selective binding energy to stabilize rate-limiting transition states or to destabilize substrates in order to lower the free energy of activation of a reaction. The realization that one can similarly direct the binding energy of an antibody molecule provided an experimental approach to test these theories, much as chemists have synthesized small molecules to test mechanistic hypotheses.

Consider, for example, the generation of an antibody that catalyses the metallation

of porphyrins. It is thought that the enzyme ferrochelatase, which catalyses this reaction in haem biosynthesis, does so by straining or distorting the planar porphyrin substrate. By generating antibodies against a synthetic mimic of a bent porphyrin, an antibody was evolved with catalytic properties similar to those of the natural enzyme. The X-ray crystal structure of the Michaelis complex provided the first direct structural support for the strain hypothesis. Similar studies have made it possible to analyse the energetic contributions of transition-state stabilization, covalent and general base catalysis, and proximity effects to biological catalysis. In some cases, such as antibodies that catalyse efficient acyl-transfer reactions, the immune system converges on the same mechanisms that are used by enzymes for a reaction, underscoring the remarkable similarity between these two evolutionary systems.

As the chemical strategies for exploiting immunological diversity have improved, so has our ability to generate efficient chemical catalysts. For example, the use of covalent immunization with a β -diketone, in which the hapten selects for an antibody with an appropriately positioned active-site lysine, gave rise to antibody aldolases with selectivities and rates that rival natural aldolases. In contrast to the corresponding enzymes, the antibodies catalyse a broad array of aldol reactions with very high enantiomeric excess.

The natural evolution of enzymes and the immunological evolution of antibodies share many features. Both involve recombination, point mutation and selection. Because the last of these can occur on a laboratory timescale, it provides an opportunity to examine the evolution of binding energy and catalysis in nature. For example, structural and biophysical studies of the immunological evolution of four catalytic antibodies (catalysing acyl transfer, oxidative, pericyclic and metallation reactions) showed that high-affinity catalytic antibodies bind to their ligands in accordance with Fischer's 'lock and key' model. In contrast, the active sites of low-affinity germline precursors undergo significant conformational changes upon ligand binding, resulting in increased antibody–ligand complementarity.

These structural studies of the affinity-maturation process have provided fundamental insights into the molecular origins of antibody affinity and selectivity. In particular, the early hypothesis of Felix Haurowitz and Pauling that germline antibody-combining sites can adapt to many shapes, much like a human hand, seems now to be correct. This conformational diversity, together with sequence diversity, explains the huge binding potential of the germline antibody repertoire.

Antibody catalysis

The use of immunological diversity to generate selective catalysts has come full circle, to the realization that antibodies have an intrinsic catalytic ability to destroy antigens.

These studies have also underscored the role of mutations located far from the active site in affecting protein binding and catalysis — an important lesson for protein engineering. Moreover, conformational flexibility has been shown to play a dynamic role in catalysis.

In chemistry, we are seeing an increased focus on the synthesis of molecules with defined biological, chemical or material properties. Unfortunately, our ability to rationally design and synthesize molecules with specific functions is not as sophisticated as our ability to synthesize specific molecular structures. The demonstration that one can use chemical principles and tools (such as transition-state theory and stable transition-state analogues) to exploit immunological diversity to generate selective chemical catalysts has introduced a new strategy into synthesis. Large combinatorial libraries of molecules can now be rationally searched for a particular molecule with the desired catalytic properties. Small-molecule libraries have now been synthesized from synthetic building blocks (rather than DNA), and screened for binding affinity. This approach has even been used to search through the entire periodic table for new combinations of elements with interesting properties.

More than 100 different reactions have been catalysed by antibodies. But do endogenous catalytic antibodies exist? All antibodies have the intrinsic ability to catalyse the oxidation of water by singlet oxygen to generate hydrogen peroxide and other reactive oxygen species, endowing the antibody itself with the ability to destroy pathogens oxidatively. Thus the field of antibody catalysis has come full circle, starting with the realization that immunological diversity can be programmed to generate new chemical functions, and ending with the understanding that antibodies all along have had a remarkable catalytic role. ■

P. G. Schultz and R. A. Lerner are in the Department of Chemistry, Scripps Research Institute, La Jolla, California 92037, USA.

FURTHER READING

Haldane, J. B. S. *Enzymes* (Longmans Green, London, 1930).

Schultz, P., Yin, J. & Lerner, R. *Angew. Chem. Int. Engl. Edn* (in the press).

BRIDGEMAN ART LIBRARY



Like hands, antibodies can adopt many shapes.

Impacts in the round

John G. Spray

A well-preserved crater has been identified deep beneath the bed of the North Sea. It may well have been produced by the impact of an extraterrestrial body some 60 million years ago.

The North Sea, off Britain's eastern seaboard, has proved a rich source of oil and gas. Much time and money has been spent in surveying the area, but a different kind of payoff is described by Stewart and Allen on page 520 of this issue¹. They have revisited seismic reflection data, obtained in the early 1990s from an area about 130 kilometres offshore from the Humber estuary. From their analysis, they conclude that beneath 40 metres of sea water and a further few hundred metres of sediments lies an impact crater that is some 20 kilometres in diameter and has a multi-ring structure.

This sort of claim is important because we know so little about how impact structures are created when meteorites and comets hit planetary bodies that any new example helps. These structures are classified into two main types, simple and complex. But how they are actually created as a result of high-speed impact — collision velocities are typically 15 kilometres per second or more — is still a matter of debate. Geological processes that might mimic or later obscure the consequences of an impact further complicate interpretations.

The simplest crater is a bowl-shaped depression (Fig. 1a), like the famous Barringer Crater in Arizona. With increasing diameter, complex craters develop, successively taking the form of central peak impact structures (Fig. 1b), peak ring structures (Fig. 1c) and multi-ring basins (Fig. 1d). On Earth, multi-ring basins probably occur at diameters of 250 km or more, as suggested by the structures identified at Sudbury, Canada, and Vredefort in South Africa². On other planets they can reach sizes of 2,000 km or more (as seen, for instance, in the Orientale basin on the Moon³). So Stewart and Allen's report¹ of a putative multi-ringed impact structure of only 20 km diameter is especially intriguing.

This is one of the first detailed analyses of a probable buried impact structure obtained using three-dimensional seismic reflection studies. The Silverpit crater, as the authors name the structure, is found in sedimentary rocks of Jurassic and Cretaceous age, and is overlain by several hundred metres of later strata. This dates its time of formation to 60–65 million years ago. The geology of the southern North Sea includes a thick region of salt, of Permian age, which could provide an alternative explanation for how Silverpit formed. Such deposits can be mobilized and,

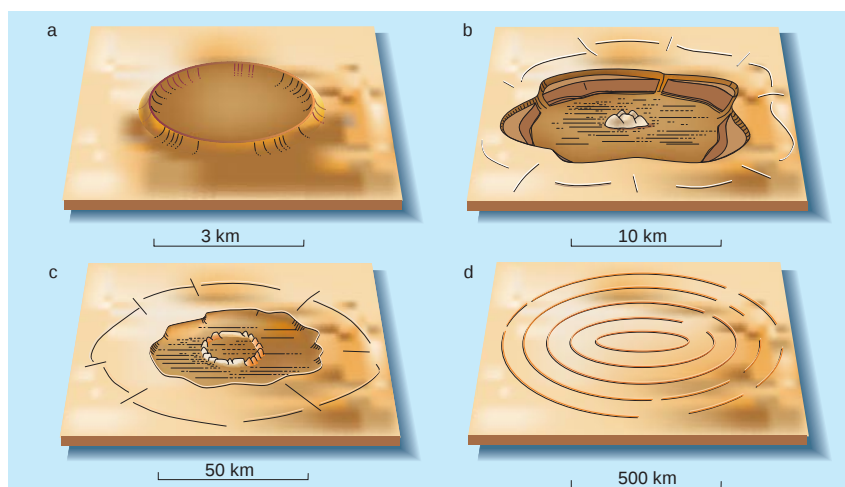


Figure 1 The main types of impact structure, as currently defined from studies of the Earth and Moon. a, At the smallest scale is the simple bowl-shaped crater. b–d, At larger scales, more complex structures are evident. b, Central peak; c, peak ring; d, multi-ring. The Silverpit crater analysed by Stewart and Allen¹ seems unusual in being a small multi-ring structure. But confirmation that the structure was indeed created by an impact will require further evidence.

because of the relatively low density of salt, can rise as finger-like vertical protrusions through the overlying strata, disrupting them. Stewart and Allen, however, rule out salt intrusion as the cause of the Silverpit crater. Their interpretation of the seismic data is that the strata beneath the crater are undisturbed, indicating that the crater was formed by forces acting from above and not from below.

The core of the Silverpit structure is a bowl-shaped crater, 3 km in diameter, with a central peak (referred to as zone 1). Numerous concentric rings surround this central crater. Zone 2 occurs 1.5–4 km from the centre and consists of a set of concentric 'fault scarps' that face the crater and form a set of tilted blocks, or steps. The faults are spaced at 300–500-m intervals and show maximum vertical displacements of 50 m. Zone 3 is 4–10 km from the centre and is defined by concentric valleys (graben, in technical terms). The fault systems in zones 2 and 3 both terminate in the lower half of the Cretaceous rock. Critically, there is some evidence that the faults in zone 2 change angle from steeper to shallower with depth and face into the centre of the structure. This makes them 'extensional' faults, and provides an explanation for the response to removal of material from zone 1 — the surrounding rocks have

moved inwards to compensate for the cavity.

Suggestions of how multi-ring structures form has been a matter of much debate. Proposed mechanisms range from gross 'Bingham fluid behaviour'⁴ (which is akin to the generation of concentric waves when a pebble is thrown into a pond) to discrete deformation by faulting of the Earth's surface⁵. If the Silverpit structure is an impact-generated feature, then faulting is clearly the ring-formation mechanism here. In this respect, the presence of 'slippery' strata beneath Silverpit may be relevant. Stewart and Allen suggest that the concentric faults may have moved at points where porous chalk is interlayered with impermeable clays. The clays could have helped to generate fluid pressures that would facilitate slip. If this was indeed the case, the multi-ringed nature of Silverpit may be a result of local geological influences and is not indicative of a new or modified class of impact structure.

More generally, Stewart and Allen's work highlights the limitations of the sample size that we have to work with in understanding impact structures on Earth. About 160 of them have been identified, but most of our knowledge is based on fieldwork on exposed examples (structures that were buried then subsequently exhumed, or were never buried). Here geologists can walk about the



100 YEARS AGO

An interesting instance of that adaptability to changing tastes and conditions which is the mainspring of progress in industry as well as in science is afforded by a note in the *Journal of the Society of Arts* (July 18). For some years the demand for claret has greatly diminished in favour of the wines of Champagne, and has seriously affected the wine industry in the Bordeaux region. Several proprietors in the Médoc have, however, now commenced the production of sparkling wines by the same process as champagne is made, and their action has been the means of developing practically a new industry. It may at first seem strange that white wine should be able to be made in the Médoc, where only black grapes are grown, but as a matter of fact champagne is almost entirely made from black grapes, and the most celebrated vineyards in the Champagne district are all planted with them... It is stated that to the ordinary taster there is nothing but the label to distinguish the sparkling médoc from the best brands of champagne.

From *Nature* 31 July 1902.

50 YEARS AGO

The Psychology of the Occult. This stimulating and highly provocative book is an attempt to describe and analyse the part "played by various types of psychological anomaly in the creation and perpetuation of occult beliefs and practices"... Mr. Rawcliffe is completely unmoved by the flood of modern propaganda in favour of the reality of so-called psychic phenomena... As for the physical phenomena of the séance room, Mr. Rawcliffe finds the evidence scarcely worth considering. To him the whole of the studies of the psychical research worker are mere examples of magic and superstition dressed up in modern garb and often presented behind a façade of statistical jargon which is intended to disguise the faulty character of the original data. In support of his position he has skilfully put together a mass of material in which the incompetence and credulity of not a few workers in this field are cruelly displayed. Yet he has omitted much that would have strengthened his case and, it must be added, a good deal that would have weakened it... however, this book remains a useful handbook for those who suspect that much of what passes for psychical research and which is often unfortunately supported by leading parapsychologists is scarcely worth the paper on which it is recorded.

From *Nature* 2 August 1952.

surface, see inside the structure and sample the rocks. But exposed impact structures are typically incomplete, lacking their uppermost parts through the vagaries of erosion and deformation processes. The only way of 'seeing' virgin morphology is when a structure has been rapidly buried after it formed, and so has been preserved. Geophysicists then need to use seismic techniques to reveal the three-dimensional structure, as has been done for Silverpit. The development of multiple concentric rings at such a small diameter may not be unusual because, until recently, we have been unable to obtain images with this degree of detail in such a well-preserved example.

Finally, the real test that Silverpit was created by an impact will be to look for shock effects in the rocks that form it. Shock-generated features such as unusual microscopic mineral deformations and shatter cones

(conical fracture systems formed by shock waves in rock) would be compelling evidence of an impact origin. Two wells that were drilled in the search for oil and gas do penetrate the structure, but unfortunately, few samples of the drill cuttings from them were taken. Given that impact structures are among the most productive hydrocarbon sites on the planet, we may get more rock samples if Silverpit shows exploration potential. ■

John G. Spray is at the Planetary and Space Science Centre, University of New Brunswick, Fredericton, New Brunswick E3B 5A3, Canada.

e-mail: jgs@unb.ca

1. Stewart, S. A. & Allen, P. J. *Nature* **418**, 520–523 (2002).
2. Grieve, R. A. F., Theriault, A. & Rupert, A. *Annu. Rev. Planet. Sci.* **28**, 305–338 (2000).
3. Spudis, P. D. *The Geology of Multi-Ring Impact Basins* 263 (Cambridge Univ. Press, 1993).
4. Baldwin, R. B. *Phys. Earth Planet. Inter.* **6**, 327–339 (1972).
5. Melosh, H. J. *Nature* **373**, 104–105 (1995).

Neurobiology

Never fear, cannabinoids are here

Pankaj Sah

Although we understand how fearful memories are stored in the brain, how they are extinguished remains a mystery. The answers may lie with the cannabinoid compounds our bodies produce.

Cannabinoids such as marijuana and hashish have been used for over a thousand years for medicinal and recreational purposes. The active 'ingredient' of these drugs is Δ^9 -tetrahydrocannabinol, which produces effects on nerve cells in the brain by binding to a protein on the neuronal surface, the CB1 receptor¹. But of course the receptor is not there simply to detect this externally derived compound: it also binds to 'endogenous' cannabinoids, which are produced naturally by the body. On page 530 of this issue, Marsicano and colleagues² propose a new role for this 'endocannabinoid' system — extinguishing fear-related memories in mice. The finding might have implications for treating anxiety disorders in humans.

We can form memories in several different ways, one of which is Pavlovian conditioning — the classic example being that of Pavlov's dogs, which learned to expect food whenever they heard a ringing tone. We all form these types of associations; for instance, we may associate a particular piece of music with our first love affair. But the connection need not always be pleasant. Imagine you are having a quiet walk in a park when you are threatened by an armed person. During the attack you are terrified; your heart races and your palms are sweaty. You run and escape. Later, you may find that entering the same park brings back in detail the memory of the attack, right down to the sweaty palms.

In the lab, the neuronal and molecular mechanisms underlying fearful memories are

often studied in animals by using 'fear conditioning'. Here, a neutral — or conditioned — stimulus, which is typically a tone or a light, is paired with an aversive (unconditioned) stimulus, typically a small electric shock to the foot. After the two stimuli are paired a few times, the conditioned stimulus alone evokes the stereotypical features of the fearful response to the unconditioned stimulus, including changes in heart rate and blood pressure and freezing of ongoing movements. Repeated presentation of the conditioned stimulus alone leads to extinction of the fearful response — the animal learns that it need no longer fear a shock from the tone or light.

A large body of work has established that a small, almond-shaped region in the brain, the amygdala, is crucial in acquiring and, possibly, storing the memory of conditioned fear^{3,4}. It is thought that, at the cellular and molecular level, this learned behaviour requires neurons in the basolateral part of the amygdala, and changes in the strength of their connection with other neurons ('synaptic plasticity') that depend on the NMDA receptor⁵, which responds to the neurotransmitter glutamate.

The extinction of aversive memories also involves the basolateral amygdala, but the cellular and molecular details are less clear. Infusing antagonists of the NMDA receptor into this region blocks extinction, implying that these receptors are important here, too⁶. Yet their exact role is not known. It has been proposed that synaptic plasticity is

again involved⁶, but the possible sites of plasticity and the underlying physiology are not known, and NMDA-receptor-dependent plasticity has not yet been correlated with extinction. Moreover, it has been suggested that there are also NMDA-receptor-independent mechanisms of extinction⁷.

Marsicano *et al.*² now propose just such a mechanism, which involves the endocannabinoids anandamide and 2-arachidonylglycerol, and their CB1 receptors. These receptors are some of the most abundant neuromodulatory receptors in the central nervous system and are expressed at high levels in the limbic system, cerebellum and basal ganglia⁸. The classical behavioural effects of exogenous cannabinoids — such as sedation and memory changes — have been correlated with the presence of CB1 receptors in the limbic system and striatum.

It has been difficult, however, to pin down the physiological role of endocannabinoids and how they are released in these regions. In studies that were the first to reveal such a role, the depolarization of neurons by repetitive activity led to the release of endocannabinoids⁹, which diffused to the terminals of other neurons and inhibited neurotransmitter release. This effect was transient in the hippocampus and cerebellum⁹ and long lasting in the striatum¹⁰. Yet these changes in neurotransmission have not been connected to any specific behavioural effects. So the study by Marsicano *et al.*² represents a leap forward in two areas of neurobiology, in that it clearly implicates the release of endocannabinoids in a well-known, simple learning task. It also links endocannabinoid release to synaptic plasticity.

After engineering mice to lack the CB1 receptor, Marsicano *et al.* first showed that although these animals could learn and later recall the association of a tone with a foot shock, they could not extinguish the memory. A drug that antagonizes the CB1 receptor similarly prevented extinction in wild-type mice. The authors then found that during the extinction protocol (exposure to the tone alone), the levels of both anandamide and 2-arachidonylglycerol were raised in the basolateral amygdala in mutant and normal mice. This implies that a process involving activation of the CB1 receptors by endocannabinoids is essential in the extinction of conditioned fear.

Next, in experiments with slices of normal mouse brains, the authors looked at neurons in the basolateral amygdala that can release GABA (an inhibitory neurotransmitter). They found that low-frequency stimulation of these neurons leads to a long-term reduction in the release of GABA, which in turn leads to less inhibition of the connecting 'pyramidal' neurons. This long-term 'depression' — a type of synaptic plasticity — was completely blocked by the CB1-receptor

antagonist, and absent in CB1-deficient mice. These findings suggest that the endocannabinoids reduce GABA release in the basolateral amygdala, thereby helping to extinguish the fear-conditioned response. In mammals, the neurons that release GABA are largely interneurons, which can be divided into several populations on the basis of their expression of certain proteins and peptides (such as cholecystokinin). The role of endocannabinoids in reducing GABA release fits with the finding that CB1 receptors in the basolateral amygdala are present on the terminals of cholecystokinin-containing interneurons^{11,12}.

This is an entirely new cellular and molecular mechanism for extinction. But how does it tie in with the NMDA receptors? There seems little doubt that activation of these glutamate receptors in the basolateral amygdala is somehow required for extinction⁶. But Marsicano *et al.*'s brain-slice experiments were performed with blocked glutamate receptors, showing that the endocannabinoid-mediated synaptic plasticity they report does not need the NMDA receptors. So we have yet to find out how these receptors are involved in extinction.

It has been argued that the neuronal circuitry underlying fear conditioning has similarities to that responsible for fear-related clinical conditions, such as post-traumatic stress disorder⁴. Behavioural

therapies for these conditions — including systematic desensitization and imagery therapies — share features with extinction. The finding that the endocannabinoids contribute to extinction raises the possibility that drugs that target these molecules and their receptors could be useful new treatments for anxiety disorders. Finally, there is much anecdotal evidence of patients using cannabis heavily in the early stages of psychiatric illness. This has often been thought to contribute to acute illness. But it seems possible that it may instead be a form of self-medication for the sometimes extreme anxiety that these people experience. ■

Pankaj Sah is in the Division of Neuroscience, John Curtin School of Medical Research, Australian National University, Canberra, ACT 2601, Australia. e-mail: pankaj.sah@anu.edu.au

1. Matsuda, L. A. *et al.* *Nature* **346**, 561–564 (1990).
2. Marsicano, G. *et al.* *Nature* **418**, 530–534 (2002).
3. LeDoux, J. E. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
4. Davis, M. & Whalen, P. J. *Mol. Psychiatry* **6**, 13–34 (2001).
5. Blair, H. T., Schafe, G. E., Bauer, E. P., Rodrigues, S. M. & LeDoux, J. E. *Learn. Mem.* **8**, 229–242 (2001).
6. Walker, D. L. & Davis, M. *Pharmacol. Biochem. Behav.* **71**, 379–392 (2002).
7. Santini, E., Muller, R. U. & Quirk, G. J. *J. Neurosci.* **21**, 9009–9017 (2001).
8. Herkenham, M. *et al.* *J. Neurosci.* **11**, 563–583 (1991).
9. Wilson, R. I. & Nicoll, R. A. *Science* **296**, 678–682 (2002).
10. Gerde, G. L., Ronesi, J. & Lovinger, D. M. *Nature Neurosci.* **5**, 446–451 (2002).
11. Katona, I. *et al.* *J. Neurosci.* **21**, 9506–9518 (2001).
12. McDonald, A. J. & Mascagni, F. *Neuroscience* **107**, 641–652 (2001).

Earth science

Core values

John Brodholt and Francis Nimmo

Calculating the age of the Earth's solid inner core has proved to be a tricky business. But the suggestion that there is more potassium in the core than had been thought could help to reconcile differing estimates.

Potassium is a relatively insignificant element in the Earth, languishing in sixteenth place in the league table of chemical abundance. But, right now, radioactive decay of a potassium isotope, ⁴⁰K, is responsible for about 10% of the heat lost by the Earth. As ⁴⁰K has a half-life of 1.25 billion years, its decay would have produced much more heat in the past — just after the Earth formed, 4.6 billion years ago, daily decay of ⁴⁰K would have produced more heat than the present total daily heat loss of the Earth. Heat within the Earth drives processes such as convection in the mantle layer and the generation of the planet's magnetic field. So knowing how potassium and other radioactive elements are concentrated in different parts of the Earth is fundamental to understanding these processes.

There is a large concentration of potassium in the Earth's crust, and a significant proportion remains in the mantle below.

What is not known is how much is in the Earth's core. Although experimental results have been ambiguous, it has generally been thought that, because of the relatively large radius of potassium ions, not much of this element could be absorbed in the core. But new results from Gessman and Wood¹, reported in *Earth and Planetary Science Letters*, show that the amount of potassium in the core depends on the core's sulphur and oxygen content, and on the structure of the coexisting silicate melt. Their findings help to explain some of the ambiguities in the earlier data, and also enable them to estimate the maximum concentration of potassium in the Earth's core — a number that has important bearing on the age of both the inner core and the Earth's magnetic field.

Although the magnetic field is generated by fluid flow in the liquid outer core, it is generally accepted that the solid inner core also plays a fundamental role. This is because

a large part of the energy needed to drive this magnetic dynamo is thought to come either from the latent heat of crystallization of the inner core, or from convection due to the separation of a light chemical component during freezing^{2,3}.

Driving a dynamo without an inner core is extremely difficult as that would require the heat flux to be high enough to provide the necessary convection, but not so high that the core solidifies. At the July meeting on 'Study of the Earth's Deep Interior' (SEDI), D. Gubbins (Univ. Leeds) argued that it is pretty much impossible for the Earth to have established a geomagnetic dynamo without an inner core; D. Stevenson (Caltech) suggested that it might be possible to sustain a dynamo without an inner core if the core was initially superheated, perhaps by 1,000 K. But this seems unlikely, as then the core would probably have been hot enough to melt the overlying mantle.

So if an inner core is necessary for the dynamo, and given that there is clear evidence for a geomagnetic field in rocks as old as 3.5 billion years, the logical assumption is that the Earth's solid inner core must be at least that old. The problem, however, is that models of core evolution have difficulty simulating an inner core that is older than about 1 billion years⁴; the core seems to be growing too quickly. But Gessman and Wood's results¹ may help to reconcile these two incompatible views.

If the inner core has grown to its present size at a constant rate over 4.6 billion years, the heat of crystallization would provide 0.5 TW (1 TW = 10^{12} W) of the total core heat flux. Estimating the total core heat flux is difficult, but by assuming that in the lowest 250 km of the mantle heat is transported by conduction only, the answer comes out at about 7 TW. Convective upwellings from the core–mantle boundary (mantle plumes) may increase this figure to 10 TW. In the absence of internal heat sources, the difference between these estimates and the heat from crystallization must be accounted for by heat released through cooling. This would require a core-cooling rate of 120–200 K every billion years, a rate that is considered rather too high. So either the core is much younger than 4.6 billion years, or there is another internal heat source — such as radioactive heating. Using similar arguments, Labrosse *et al.*⁴ estimated that, if the inner core is as old as 4.6 billion years, around 4 TW must be released through radioactive heating.

Gessman and Wood's high-temperature, high-pressure experiments¹ show that, in the presence of sulphur and oxygen, potassium alloys with liquid iron, and that in the Earth's core a concentration of up to 250 p.p.m. could be expected. That much potassium would produce only about 1.7 TW of heat, somewhat less than the 4 TW required for an inner core whose age matches the Earth's. But it may be enough to push back the core age to about

2.5 billion years. Rama Murthy and colleagues (personal communication), however, suspect that there might be even more potassium in the core than Gessman and Wood suggest. And if the core-forming liquid equilibrated with the mantle at lower pressures than assumed by Gessman and Wood, the concentration of potassium in the core could be higher still. If the abundance of other radioactive elements, such as uranium and thorium, in the iron-rich liquid of the core is also affected by the presence of sulphur, that again would push back the age of the core.

At this point, although the experiments performed by Gessman and Wood help, they do not fully resolve the problem of the age of the Earth's inner core and dynamo. If the core really can store significant quantities of potassium, geochemical arguments regarding stratification of the mantle will also have to be revisited⁵. Clearly, more theoretical and

experimental work is needed. It is worth noting, however, that Gessman and Wood's results apply to other planetary cores. The inner core of Mars formed at lower pressure and is thought to have a high sulphur content, so it could also have very high concentrations of potassium: radioactive decay may have provided the energy for an early magnetic dynamo on Mars.

John Brodholt and Francis Nimmo are in the Department of Earth Science, University College London, Gower Street, London WC1E 6BT, UK. e-mail: j.brodholt@ucl.ac.uk

1. Gessman, C. & Wood, B. J. *Earth Planet. Sci. Lett.* **200**, 63–78 (2002).
2. Gubbins, D., Masters, T. G. & Jacobs, J. A. *Geophys. J. R. Astron. Soc.* **59**, 57–99 (1979).
3. Buffett, B. A. *Geophys. Res. Lett.* **29**, 10.1029/2001GL014649 (2002).
4. Labrosse, S., Poirier, J.-P. & Le Mouél, J.-L. *Earth Planet. Sci. Lett.* **190**, 111–123 (2001).
5. Kellogg, L. H., Hager, B. H. & van der Hilst, R. *Science* **283**, 1881–1884 (1999).

Materials

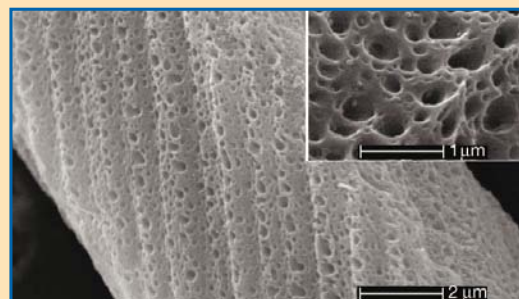
Zeolites branch out

Wood, with applications ranging from construction to cabinet-making, has just notched up another use. As they describe in *Advanced Materials*, Angang Dong and colleagues (*Adv. Mater.* **14**, 926–929; 2002) have used it as a template for creating zeolites.

Zeolites are porous structures of crystalline aluminosilicate, made on templates that define the size of the pores and their interconnectivity. They can be designed for specific applications; for example, the pore size can be chosen to set the rate at which other molecules will diffuse through the zeolitic structure.

Templates such as shells have already been used to synthesize inorganic analogues of biological materials. As the structural organization of such materials cannot be finely controlled, some of the advantages of rationally designed synthetic templates are lost. But other properties make biological templates attractive: they can be cheap, abundant, renewable, and so commercially viable; and they are also environmentally benign.

Dong *et al.* report the synthesis of a zeolitic material possessing a porosity inherited



from its wood templates — bamboo (a grass with woody stems) and cedar (a conifer). Further advantages of wood are its intrinsically complex and hierarchical pore structure, and the abundance of tree species — and so morphologies.

To make a zeolite, the authors used a 'seeded growth' technique. First they treated the wood templates with a polyelectrolyte, and dispersed pre-formed zeolite nanocrystals through the templates. Then they washed the nanocrystal-coated wood slices and immersed them in zeolite solution for a day. The zeolite–wood composite was heated in air to remove the wood, leaving a free-standing zeolite composed of thin, uniform, continuous membranes.

Using X-ray diffraction on both the bamboo and the cedar templates, Dong *et al.* verified

that the resulting materials were purely zeolitic. Scanning-electron-microscope images revealed that the synthetic strategy had not destroyed the skeleton of the original tissue. Instead, the zeolite nanocrystals had completely penetrated the wood, faithfully replicating the wood's cellular structure in fine detail. The cedar-template zeolite (shown above) was composed of interconnected bundles of micrometre-scale hollow vessels, with some fibres having features such as pits (which originally connected adjacent cells) and helical stripes. The bamboo zeolite showed rings and spine-like structures.

Applications of zeolites formed from wood templates are still some way off. But Dong *et al.* predict that they could ultimately be useful in catalytic and adsorption technologies. **Rosamund Daw**

Making progress with limb models

Denis Duboule

What is the developmental mechanism that makes our upper arms different from our forearms or fingers? Two new papers challenge an influential and popular model, and propose an alternative view.

Our limbs can move freely because they are constructed in segments, separated by articulations. The arm, for example, is composed of the upper arm, the forearm and the hand — three parts that develop from an early bud of unspecialized limb cells. How do these cells 'decide' which parts of the limb to form? Signals coming from a tissue in the limb bud called the apical ectodermal ridge (AER)¹ are known to be important in limb development, and in one long-standing model, these signals instruct cells of the upper arm to form first, followed by cells of the forearm, and finally cells of the hand. But on pages 539 and 501 of this issue, Dudley and colleagues² and Sun and co-workers³ put forward a different view. They propose that all these cell types are produced at the same time in the limb bud, and then — under the control of the AER — the regions of different cells expand at different times to form the complete limb.

This aspect of limb development is called proximal-to-distal patterning (in our example, 'distal' refers to the hand and 'proximal' to the upper arm), and it has been extensively studied in birds because their wing buds are highly amenable to experimental manipulation. Such studies revealed, for instance, the importance of the AER — a layer of tissue that covers the rim of the distal tip of limb buds¹. Removing the AER early in development led to loss of most of the wing, whereas on late removal only the digits were absent^{4,5}.

Until now these results have been explained by an influential view of proximal-to-distal patterning called the 'progress zone' model⁶. Here, an internal clock — controlled by the AER — dictates the fate of a population of cells located in the progress zone of the limb bud, just underneath the AER. The idea is that the fate of these cells is not fixed from the start, but depends on how long they are in this zone and hence under the influence of the AER. As the cells proliferate, they leave the progress zone and escape the control of the AER. Cells that leave early contribute to proximal limb regions. Cells that leave late have been under the control of the AER for longer, and adopt a distal fate (Fig. 1a). This model can explain the results of AER removal^{4,5}; for instance, if the AER is deleted later in development, proximal regions can form but cells in the progress zone have not been under the influence of the AER for long enough to make distal regions (Fig. 1c).

This conceptual framework can accom-

modate many other experimental observations. But the new results^{2,3} do not easily fit with this view. Both new papers suggest that positional information is not acquired in a time sequence, but instead is already dispatched in the early limb bud. In this model, it is the selective multiplication of prespecified cohorts of cells that is time-dependent, with proximal cells proliferating earlier (Fig. 1b).

Dudley *et al.*² started by reassessing how cells behave at the tip of the wing bud immediately after the AER is removed, as it had previously been observed⁷ that some cells die soon after AER removal. Dudley *et al.* confirm this finding, and also show that cell proliferation drops off dramatically. The domain of cell loss is always the same size, regardless of when the AER is removed. So if the AER is ablated late, more of the limb has had the chance to develop, and a smaller proportion of the limb (only the distal region, which is right under the AER) is affected (Fig. 1d). In other words, this view can

also explain the effects^{4,5} of AER removal.

The authors went on to label cells in early limb buds and to look at their progeny at later stages. As expected, cells labelled within the distal tip of the bud colonized distal skeletal elements. After AER removal, however, these cells disappeared. This does not fit with the progress-zone model, which would predict that these cells would simply be incorporated into more proximal elements (because they have been under the influence of the AER for less time). So Dudley *et al.* again conclude that these cells died, and suggest that the AER helps to prevent cell death and maintain proliferation — not to control cell fate — in the underlying 200- μ m tissue layer.

Dudley *et al.* also propose that cells in the limb bud are specified as 'proximal' or 'distal' early on — a claim substantiated by labelling cells in normal budding limbs². For example, when the authors labelled cells in the distal part of the limb bud, the signal was subsequently found throughout the region that formed digits. Labelled cells were rarely found in two adjacent territories, suggesting that precursor cells for all segments coexist in the early bud. That is not to say, however, that the destiny of these cells becomes fixed at the same time: precursors intended for the digits could still change their fate later than the forearm cells, and long after the upper arm cells, could do so.

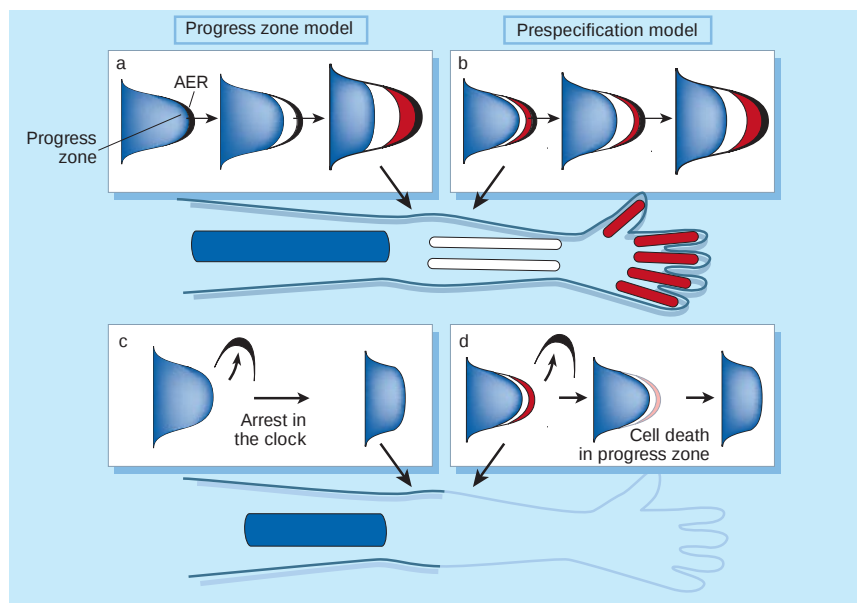


Figure 1 Developing limbs. a, b, Two views of proximal-to-distal patterning in developing limb buds. a, In the 'progress zone' model, the fate of proximal (prospective upper arm; blue) elements is specified before the fate of more distal elements, as the bud grows. Changes in cell fate occur in the progress zone, just next to the apical ectodermal ridge (AER). b, In the prespecification model proposed by Dudley *et al.*² and Sun *et al.*³, proximal-to-distal fates are already specified early on and the observed time sequence in skeletal development results from the selective expansion of these prespecified domains, along with cellular determination — the acquisition of definitive cell fate. c, d, How the two models can explain the effects^{4,5} of removing the AER. c, According to the progress zone model, the specification clock is arrested and (in this case) distal specification never occurs. d, In the prespecification model, the expansion of prespecified distal domains is aborted because of cell death. In both cases, the same results are expected.

The AER exerts its effects through the release and diffusion of members of the fibroblast growth factor (FGF) protein family, particularly through FGF4 and FGF8, which can substitute for the ridge if applied to the limb bud's distal tip^{8,9}. To find out exactly what FGF4 and FGF8 do, Sun *et al.*³ set out to produce mice lacking both proteins in only the limb buds. They conclude that an AER lacking both FGF4 and FGF8 cannot promote limb development, even though other FGFs remain.

Interestingly, however, in forelimb buds the genes for FGF4 and FGF8 were initially active. Therefore, protein deletion was slightly delayed, so that a burst of FGF8 still occurred in the nascent AER and in turn activated the FGF4 gene. Both genes were inactivated soon after, and neither protein was detected after budding. Nonetheless, this short exposure to FGF4 and FGF8 was enough to allow some skeletal elements to develop, in particular the humerus (in the upper forelimb). Unexpectedly, in most cases more distal elements (albeit much shorter than normal) were also identified.

From their analysis of the mutant forelimb and hindlimb buds, Sun *et al.* conclude that FGF signalling from the AER has two important functions. First, it is necessary at the earliest stages of limb development to establish a limb bud of normal size. Later, it is needed for limb-bud cells to survive. The upshot is that, without FGF4 and FGF8, there are not enough cells in the growing bud to produce the required skeletal elements.

The pattern of skeletal elements that formed in the mutant forelimbs, including distal (albeit abnormal) bones, is hard to reconcile with the progress zone model, which would have predicted the presence of normal proximal bones and no distal bones. Instead, what was observed³ fits with the view expressed by Dudley *et al.*² that cells that are the precursors of all limb elements exist together in the limb bud and then multiply and survive under the influence of proteins produced by the AER. If this is so, it should be possible to find marker genes for the different cells in the early limb bud.

Does this mean that the progress zone model should be dismissed, and the textbooks rewritten? Not quite. Some of the new results might be explained by a loose interpretation of this model, as the authors acknowledge, and some pieces of the new theory are still missing. It is nonetheless clear that another, equally substantiated proposal has entered the game. As Lewis Wolpert has said, "What you see today happened yesterday"; perhaps this wise statement should also be applied to limb patterning. Whatever the case, models exist to be challenged, and there is little doubt that this confrontation of ideas will significantly advance the field. ■

Denis Duboule is in the Department of Zoology and Animal Biology, University of Geneva, Sciences III,

Quai Ernest Ansermet 30, 1211 Geneva 4, Switzerland.
e-mail: Denis.Duboule@zoo.unige.ch

1. Rubin, L. & Saunders, J. W. *Dev. Biol.* **28**, 94–112 (1972).
2. Dudley, A. T., Ros, M. A. & Tabin, C. J. *Nature* **418**, 539–544 (2002).
3. Sun, X., Mariani, F. V. & Martin, G. R. *Nature* **418**, 501–508 (2002).
4. Saunders, J. W. *J. Exp. Zool.* **108**, 363–403 (1948).

5. Summerbell, D. J. *Embryol. Exp. Morph.* **32**, 651–660 (1974).
6. Summerbell, D., Lewis, J. H. & Wolpert, L. *Nature* **244**, 492–496 (1973).
7. Rowe, D. A., Cairns, J. M. & Fallon, J. F. *Dev. Biol.* **93**, 83–91 (1982).
8. Niswander, L., Tickle, C., Vogel, A., Booth, I. & Martin, G. R. *Cell* **75**, 579–587 (1993).
9. Fallon, J. F. *et al. Science* **264**, 104–107 (1994).

Applied physics

Ultrafast magnetic switching

Burkard Hillebrands and Jürgen Fassbender

Magnetic-memory devices rely on fast switching of the magnetization vector in a material to store data. The speed of such devices could be increased by adding a magnetic-pulse-shaping technique.

The demand for speed in data processing must be matched by the fast response of memory devices. One promising path is the realization of magnetic random access memory (MRAM): a logical bit is stored in a unit cell of magnetic material by setting the orientation of the magnetization vector inside the cell in one of two possible directions, coding the logical values '0' and '1'. To write data, an external magnetic field is applied to reverse the magnetization, flipping the bit between '0' and '1'. But the speed, and the reliability, with which bits can be set is still limited. However, on page 509 of this issue, Gerrits *et al.*¹ report a new approach by which ultrafast magnetization reversal can be achieved.

Conventionally, the bit state has been changed by applying an external magnetic field, which causes nucleation of magnetic domains in the unit cell followed by expansion of the magnetic domains. The speed of both processes is inherently limited. But it was realized that if an instantaneous magnetic-field pulse was used instead, reversal could happen more quickly by 'precession'. Back *et al.*^{2,3} were the first to demonstrate this effect, indirectly, by using a pulsed beam of relativistic electrons to generate a magnetic-field pulse in a thin film of cobalt (a metal that, like iron, has strong magnetic properties). The field pulse rotates the magnetization vector out of the film plane, inducing a 'demagnetizing' field perpendicular to the plane. After the field pulse is gone, the magnetization vector precesses — like a child's spinning-top — around the demagnetizing-field direction before settling into the switched state.

Although precession is required during the switching process, it is a nuisance after switching is complete. Unwanted precession (or 'magnetic ringing') could be minimized if the initial and final positions of the magnetization direction were connected by the path of just half a precession loop. Shaping the magnetic-field pulse, as has been predicted^{4,5}, could be a way to achieve this, and several groups have been working to realize this switching mechanism in the laboratory in a device-oriented approach. Here mag-

netic-field pulses are generated by electric-current pulses launched into microstrip lines beneath the magnetic cells.

Two groups^{6,7} have reported the suppression of magnetic ringing in the case where the magnetization deviates from the original direction but is not switched into a new equilibrium direction. This ringing suppression occurs, for instance, if a short, weak magnetic-field pulse is applied — if the pulse is terminated exactly at the point when the magnetization is parallel to the effective field direction, the magnetization direction stops moving and ringing is avoided. This effect has been reproduced previously by Gerrits *et al.*⁸, and by Schumacher *et al.*⁹ for technologically relevant giant-magnetoresistance devices.

So then the competition was open for the first demonstration of precessional switching using magnetic-field pulses generated by electric-current pulses. And at the 'Magnetism and Magnetic Materials' (MMM) conference held in November 2001, the first successful experiments were reported independently by three groups^{10,11}, including Gerrits *et al.* whose detailed results are reported here¹.

By quantitatively measuring the magnetization trajectory in a micrometre-size thin-film element of ferromagnetic permalloy, Gerrits *et al.* have demonstrated switching with suppressed ringing using a specially shaped magnetic-field pulse. They sent two laser pulses, with a time delay between them, onto two semiconductor light-sensitive switches (Austin switches), generating two electric-current pulses. Because an electric current always induces an accompanying magnetic field, superimposing the two current pulses led to the creation of a shaped magnetic-field pulse. Using sophisticated magneto-optic techniques, the authors were able to track the effect of the shaped pulse on magnetization switching at five different positions in the thin-film cell, and showed that the magnetization rotation and ringing suppression is coherent throughout the entire cell.

One particular property of precessional switching is that both switching directions, from '0' to '1' and from '1' to '0', can be achieved (with certain geometry) by apply-

ing the same, unipolar pulse pointing along the hard axis of the magnetization — the magnetization direction is 'toggled' by each pulse. This effect was reported at the MMM conference by Schumacher *et al.*¹¹ in their work with giant-magnetoresistance devices, which demonstrated multiple precessional switching as the pulse length was varied.

Gerrits and colleagues' achieved switching time of 2×10^{-10} s implies an impressive data-writing rate of 5 GHz. But they point out that optimizing the material parameters as well as the field-pulse shape could shorten the write-time still further. Their verification of the concept of precessional switching through field-pulse adjustment is a step towards the future of ultrafast switching in devices. ■

Burkard Hillebrands and Jürgen Fassbender are in

the Fachbereich Physik, Universität Kaiserslautern, Erwin-Schrödinger-Straße 56, 67663 Kaiserslautern, Germany.
e-mails: hilleb@physik.uni-kl.de
fassbend@physik.uni-kl.de

1. Gerrits, Th. *et al.* *Nature* **418**, 509–512 (2002).
2. Back, C. H. *et al.* *Phys. Rev. Lett.* **81**, 3251–3254 (1998).
3. Back, C. H. *et al.* *Science* **285**, 864–867 (1999).
4. Bauer, M., Fassbender, J., Hillebrands, B. & Stamps, R. L. *Phys. Rev. B* **61**, 3410–3416 (2000).
5. Miltat, J., Albuquerque, G. & Thiaville, A. in *Spin Dynamics in Confined Magnetic Structures* (eds Hillebrands, B. & Ounadjela, K.) 1–33 (Springer, Berlin, 2002).
6. Crawford, T. M., Kabos, P. & Silva, T. J. *Appl. Phys. Lett.* **76**, 2113–2115 (2000).
7. Bauer, M., Lopusnik, R., Fassbender, J. & Hillebrands, B. *Appl. Phys. Lett.* **76**, 2758–2760 (2000).
8. Gerrits, Th. *et al.* *J. Magn. Magn. Mater.* **240**, 283–286 (2002).
9. Schumacher, H. W. *et al.* *Appl. Phys. Lett.* **80**, 3781–3783 (2002).
10. Kaka, S. & Russek, S. E. *Appl. Phys. Lett.* **80**, 2958–2960 (2002).
11. Schumacher, H. W. *et al.* *IEEE Trans. Mag.* (in the press).

Cell cycle

Oscillation sensation

David O. Morgan and James M. Roberts

Various molecular actors play out the scene in which a cell exits mitosis — nuclear division — before entering the next cell cycle. One of them, the protein Clb2, steps back into the spotlight. Another, Clb5, bows out.

In 1982, while studying events in the early sea-urchin embryo, Tim Hunt¹ found that one particular protein — which he called cyclin — is made and destroyed in each cell cycle. Murray and Kirschner² later discovered that, in frog embryos, cyclins are the only proteins that need to be newly synthesized for the cell cycle to occur. Frog eggs are giant storehouses of many proteins necessary for cell division. So why couldn't cyclins be stockpiled like the others? The answer came from a key experiment by the same investigators³: it showed that cyclin degradation can be prevented by removing a 'destruction box' sequence on the protein, and with dramatic results. When cyclin turnover was prevented, the cell cycle was blocked in mitosis. In other words, the alternating synthesis and degradation of cyclin is necessary for entry into and exit from mitosis. Here was a first glimpse of the molecular oscillator that underlies the cyclical nature of cell division. On page 556 of this issue⁴, Wäsch and Cross bring our understanding of this oscillator closer to maturity.

The cell-cycle system is stripped to its essentials in early metazoan embryos such as those of the frog *Xenopus laevis* (Fig. 1a). Here, the cycle simply alternates between the S (DNA replication) and M (mitotic) phases. The levels of mitotic cyclins rise gradually as the cell approaches mitosis, leading to a peak of activity of the enzyme cyclin-dependent kinase (Cdk) that propels the cell cycle to metaphase — the phase in which the duplicated chromosomes are aligned on the mitotic spindle. Cyclin is then abruptly

destroyed at the transition from metaphase to the next phase, anaphase, when the chromosomes separate irreversibly into two daughter sets of chromosomes. The levels of cyclin and its associated Cdk activity drop to negligible levels, which is required for the cell to disassemble the mitotic spindle and exit mitosis.

Oscillations in cyclin abundance, and hence in Cdk activity, occur because Cdk can turn themselves off. This classic example of a negative-feedback oscillator includes two basic components: an activator, the cyclin–Cdk complex, and a repressor called the anaphase-promoting complex (APC)^{5–7}

(Fig. 1a). The APC ubiquitinates proteins, thereby initiating their degradation, and is activated by a subunit, Cdc20, that recruits substrates such as the cyclins. A central feature of this system is that Cdks directly phosphorylate and thereby promote activation of Cdc20–APC⁸. The rise in Cdk activity in mitosis therefore leads, with some necessary delay, to activation of Cdc20–APC, which triggers cyclin destruction. The resulting drop in Cdk activity resets the system, allowing the APC to return to an inactive state. Cyclins begin to accumulate again and the cycle begins anew.

Because the early embryonic cell cycle has only S and M phases, it can be managed by this simple oscillator. The cell cycle in somatic cells and yeast, however, includes a G1 phase, in which the cell can pause to allow growth or differentiation (Fig. 1b). Inclusion of G1 is achieved by a second oscillator that also degrades mitotic cyclins and thereby stably suppresses Cdk activity. But it differs from the first oscillator in using the APC-activating subunit Cdh1 instead of Cdc20 (Fig. 1b). It is also controlled differently: whereas Cdc20–APC is activated by Cdks, Cdh1–APC is inhibited by them. In late mitosis, therefore, inactivation of Cdks by Cdc20–APC causes activation of Cdh1–APC. In this way, cyclin destruction and Cdk inactivity persist after mitosis, creating a stable G1 state. This state is maintained until the Cdh1–APC is turned off at the G1/S boundary by G1 cyclin–Cdks that are not substrates of the APC.

But do similar regulatory systems operate in all eukaryotes? Work on the budding yeast *Saccharomyces cerevisiae* suggested that in this species Cdc20 and Cdh1 might not fit so neatly into models developed in frogs and other metazoans. Several studies concluded that, at least in budding yeast, Cdc20–APC is not directly responsible for triggering Cdk inactivation and so mitotic exit. Instead, it

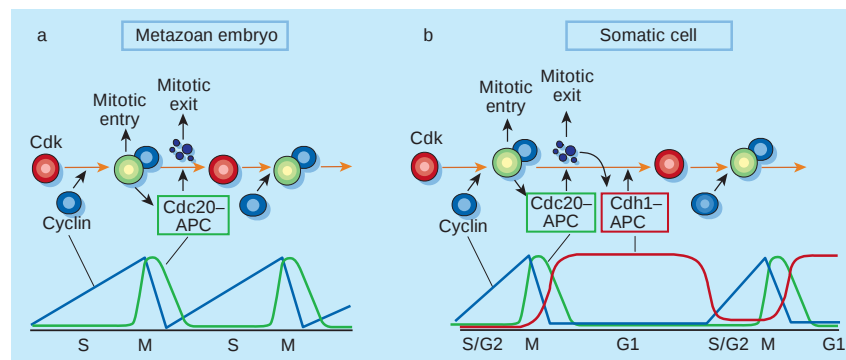


Figure 1 The cell cycle in early embryonic cells and somatic cells. **a**, In embryonic cells, which alternate between S and M phases, cyclin-dependent activation of Cdk not only leads to mitotic entry but also promotes activation of Cdc20–APC, which then triggers cyclin destruction and mitotic exit. The loss of Cdk activity causes Cdc20–APC inactivation, and cyclin can accumulate again. **b**, Somatic cells have an additional G1 phase. Cdk inactivation in late mitosis results in the activation of Cdh1–APC, which maintains cyclin destruction (and so Cdk inactivity) throughout G1. Phases of the cell cycle are S ('synthesis', in which DNA replication occurs) and M (nuclear division, or mitosis, followed by cell division, or cytokinesis). G1 is a 'gap' phase in which growth takes place.

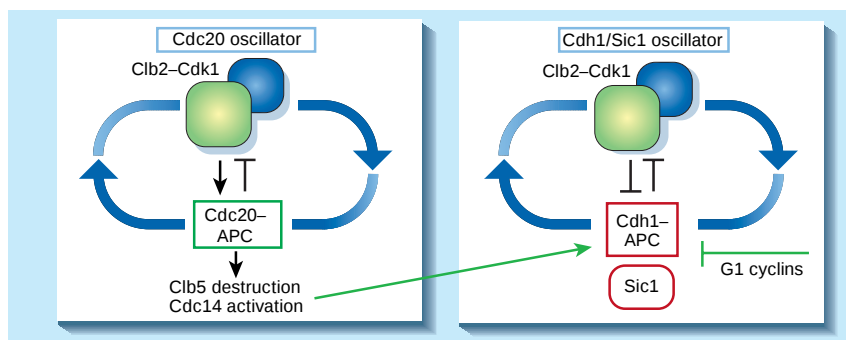


Figure 2 Cell-cycle control in budding yeast. The implication of Wäsch and Cross's results⁴ is that in yeast — and possibly all eukaryotic cells — control may be exercised by two complementary oscillators. Blue arrows show the movements of the oscillators between a state of Cdk activity (which drives mitotic entry) and a mutually exclusive state of APC activity (which drives mitotic exit). Wäsch and Cross show that the Cdc20-based oscillator of yeast is similar to that of metazoan embryos (Fig. 1a). Mitotic Cdk activity (Clb2–Cdk1) promotes Cdc20–APC activation, which then triggers sufficient Clb2 destruction for mitotic exit. The Cdh1/Sic1 oscillator requires extrinsic inputs (green arrows) to switch between the mutually inhibitory Cdk and APC active states. In wild-type yeast cells, Cdc20–APC promotes activation of Cdh1–APC and Sic1 by causing destruction of Clb5 (and Pds1, whose destruction may initiate Cdc14 activation^{13,14}). G1 cyclins turn off the Cdh1–APC.

was proposed that destruction of the major mitotic cyclin, Clb2, was primarily the job of the other APC regulator, Cdh1^{9,10}. This raised the issue of whether the regulation of mitosis in metazoan embryos could guide understanding of the somatic cell cycle.

If yeast Cdc20 does not directly target mitotic cyclins for degradation, then what does it do? One function is to promote the destruction of a protein called Pds1, which results in sister chromatid separation. Thus, cells lacking Cdc20 (*cdc20Δ*) arrest in metaphase before sister separation. Cells lacking both Cdc20 and Pds1 (*cdc20Δ pds1Δ*) undergo separation of sister chromatids, indicating that Pds1 is the Cdc20 target that prevents the onset of anaphase. Nevertheless, these cells arrest in anaphase with high Cdk activity, suggesting that at least one more Cdc20 target in addition to Pds1 must be destroyed to allow mitotic exit.

To identify this target, members of the Nasmyth laboratory looked for genes whose deletion would allow *cdc20Δ pds1Δ* cells to survive¹¹. The expectation from the work with frog embryos and other systems was that the second APC target would be Clb2. However, the gene identified in this work was *CLB5*, which encodes a cyclin involved in initiating DNA replication. Clb5 was found to be a target of Cdc20–APC. The striking ability of *cdc20Δ pds1Δ clb5Δ* cells to survive led to the idea that the Cdc20-dependent destruction of Clb5 is required for mitotic exit. It was proposed that a drop in Clb5–Cdk activity turns on Cdh1, resulting in the destruction of Clb2.

At this point, two more regulators in yeast must be mentioned^{5,6}. One is a Cdk inhibitor protein, Sic1. Like Cdh1, Sic1 is inhibited by direct Cdk phosphorylation; Cdk inactivation in late mitosis therefore leads to activation of Sic1, which further dampens Cdk activity in M/G1. The other regulator is

Cdc14, which is essential for mitotic exit. It functions, at least in part, by dephosphorylating (and therefore activating) Cdh1 and Sic1.

Wäsch and Cross⁴ now enter the scene. Their results not only raise doubts about the importance of Clb5 destruction in mitotic exit, but highlight the role of Cdc20–APC in destroying Clb2. They tested the conclusion of the Nasmyth group by revisiting Murray and Kirschner's experiment, but in yeast with Clb5; if Clb5 degradation is necessary for mitotic exit then a stable mutant of Clb5 should block mitosis. Wäsch and Cross constructed a yeast strain (*CLB5Δdb*) in which they replaced the *CLB5* gene with one that encodes a version lacking the destruction-box sequence that targets Clb5 to Cdc20–APC. Surprisingly, these cells had no defect in mitotic exit, implying that destruction of Clb5 is not required. Mitotic exit in *CLB5Δdb* cells was not due to inhibition of Clb5–Cdk activity by Sic1: exit also occurred in *CLB5Δdb* cells lacking Sic1.

If Clb5 is not the key Cdc20 target whose destruction promotes exit, then what is? Wäsch and Cross looked again at Cdc20-dependent turnover of Clb2 (which for technical reasons had not been analysed by the Nasmyth group). Indeed, Surana and colleagues had reported some Cdc20-dependent destruction of Clb2 in anaphase cells¹². In a parallel experiment to that with Clb5, Wäsch and Cross constructed a *CLB2Δdb* yeast strain. Unlike *CLB5Δdb* cells, *CLB2Δdb* cells were unable to exit mitosis (the strain could only be propagated by conditional overexpression of the Cdk inhibitor Sic1). Additional experiments showed that the destruction of Clb2 was at least partly dependent on Cdc20. In total, these results support the idea that Cdc20–APC activation leads to partial destruction of Clb2, and that this is necessary for mitotic exit.

Wäsch and Cross also explored whether Cdc20–APC alone is sufficient to allow mitotic exit, by studying cells lacking Cdh1 and Sic1. These studies are particularly provocative because it was known that cells lacking these proteins are not viable, suggesting that mitotic exit could not occur in their absence. They found, however, that cells lacking both Cdh1 and Sic1 do not have a pronounced defect in mitotic exit: the cells can disassemble their mitotic spindles, albeit with some delay and occasional abnormalities. Entry into G1 is not completely normal, however, because of defects in DNA replication. Cdk inactivation is required for resetting of replication origins in G1, and incomplete Cdk inactivation in these cells may cause lethal replication defects.

Wäsch and Cross's results suggest that yeast cells operate with two distinct mitotic oscillators, based on Cdc20 and Cdh1/Sic1 (Fig. 2). Normally, the Cdc20–APC system directly initiates mitotic exit by promoting partial Clb2 destruction. So the core of the yeast cell-cycle system has the same molecular circuitry as the 'simple' cell cycles of metazoan embryos. Neither oscillator is strictly required for the cell cycle, because under artificial conditions just one can suffice. Yeast cells survive, albeit poorly, without Cdh1 and Sic1. On the other hand, the Cdh1/Sic1 oscillator can drive the cell cycle if Cdc20 is disabled and certain other conditions are met (Pds1 and Clb5 are also deleted).

Usually, however, both oscillators are required to ensure efficient cell-cycle progression and to provide a stable G1 state of low Cdk activity. In addition, Cdc14 is required for mitotic exit, not simply because it dephosphorylates Cdh1 and Sic1 but because it may dephosphorylate other targets involved in the process. It appears to be a component of both oscillators — further studies of which promise to unveil some of the most fundamental features of cell-cycle control. ■

David O. Morgan is in the Department of Physiology, University of California at San Francisco, San Francisco, California 94143-0444, USA.

e-mail: dmorgan@cgl.ucsf.edu

James M. Roberts is at the Howard Hughes Medical Institute, and the Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109-1024, USA.

e-mail: jroberts@fhcrc.org

- Evans, T. et al. *Cell* **33**, 389–396 (1983).
- Murray, A. W. & Kirschner, M. W. *Nature* **339**, 275–280 (1989).
- Murray, A. W., Solomon, M. & Kirschner, M. *Nature* **339**, 280–286 (1989).
- Wäsch, R. & Cross, F. *Nature* **418**, 556–562 (2002).
- Morgan, D. O. *Nature Cell Biol.* **1**, E47–E53 (1999).
- Bardin, A. J. & Amon, A. *Nature Rev. Mol. Cell Biol.* **2**, 815–826 (2001).
- Peters, J. M. *Mol. Cell* **9**, 931–943 (2002).
- Rudner, A. D. & Murray, A. W. *J. Cell Biol.* **149**, 1377–1390 (2000).
- Visintin, R., Prinz, S. & Amon, A. *Science* **278**, 460–463 (1997).
- Schwab, M., Lutum, A. S. & Seufert, W. *Cell* **90**, 683–693 (1997).
- Shirayama, M., Toth, A., Galova, M. & Nasmyth, K. *Nature* **402**, 203–207 (1999).
- Yeong, F. M. et al. *Mol. Cell* **5**, 501–511 (2000).
- Tinker-Kulberg, R. & Morgan, D. O. *Genes Dev.* **13**, 1936–1949 (1999).
- Stegmeier, F., Visintin, R. & Amon, A. *Cell* **108**, 207–220 (2002).

Maturation of mouse fetal germ cells *in vitro*

Even immature oocytes can eventually be fertilized after some skilful manipulation.

Nuclear reprogramming is essential during gametogenesis for the production of totipotent zygotes. Here we show that premeiotic female germ cells derived from mouse fetuses as early as 12.5 days post coitum are able to complete meiosis and genomic imprinting *in vitro* and that these matured oocytes are highly competent in supporting development to full term after nuclear transfer and *in vitro* fertilization. To our knowledge, this is the first time that complete oogenesis has been successfully accomplished *in vitro*.

Although the ovaries of mammals contain thousands or millions of immature oocytes, few of these ever mature to the point at which reproduction *in vivo* is possible. Ovarian oocytes therefore constitute a large and potentially valuable resource for clinical and zoological application. However, although mature oocytes have been produced *in vitro* by culturing immature oocytes^{1–3}, oogenesis was never complete, and attempts to produce even non-growing oocytes at the diplotene stage of the first meiosis from ovaries derived from newborn mice have met with limited success (only 0.016%; ref. 2).

We have shown that this poor ability of oocytes to mature in culture is due to their incompetent cytoplasm⁴, a problem that can be overcome by transferring their nuclei into enucleated, fully grown oocytes. Although live pups can eventually be produced from oocytes reconstituted in this way, this is not possible if nuclei from small, immature oocytes are used for reconstitution, probably because of defects in their meiotic chromosomal configuration and/or genomic imprinting^{5–7}.

We cultured premeiotic female germ cells *in vitro* in an attempt to complete this essential nuclear reprogramming (see supplementary information for details). Ovaries from mouse fetuses at 12.5 days post coitum (d.p.c.), with the mesonephroi attached, were cultured for 7 days, followed by removal of the mesonephroi and a further 10 days in culture (Fig. 1). These cultured ovaries contained many secondary follicles, which we isolated and cultured for 11 days; at the end of the 28-day culture period, some follicles showed antrum formation and the oocytes had increased in diameter (63.9 μm , $n = 127$; Fig. 1).

We estimated the extent of genomic imprinting in these oocytes by analysing DNA methylation in the imprinted gene *Igf2r* (ref. 8). The methylation pattern at each stage of culture was consistent with that in oocytes taken from mice at the

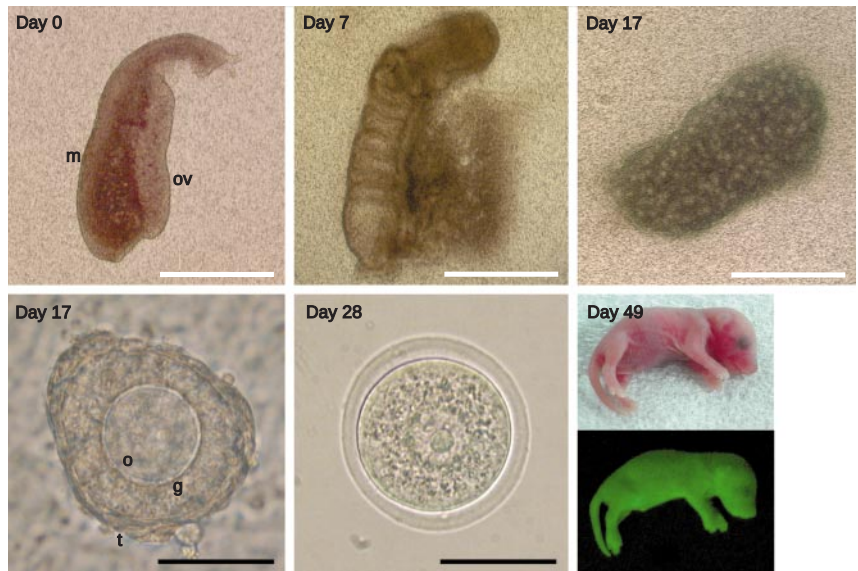


Figure 1 Development of premeiotic female germ cells *in vitro*. Day 0, ovary (ov) from a green-fluorescent-protein (GFP)-labelled B6CBF1 (GFP-C57BL/6J $\times$ CBA) transgenic mouse fetus at 12.5 d.p.c.¹¹, with attached mesonephroi (m). Ovaries were cultured in Waymouth medium supplemented with 10% fetal bovine serum (FBS) on a Costar Transwell membrane². At day 7 the mesonephroi was removed and at day 17 a secondary follicle, consisting of theca cells (t) and 2–3 layers of granulosa cells (g) around the oocyte (o), was isolated from the ovary and cultured in MEM- α medium supplemented with 5% FBS, 0.1 IU ml⁻¹ follicle-stimulating hormone, 5 μg ml⁻¹ insulin, 5 μg ml⁻¹ transferrin and 5 ng ml⁻¹ selenium on a Costar Transwell-COL membrane^{2,3}. Day 28, oocyte isolated from cultured follicle for nuclear transfer. After nuclear transfer and *in vitro* fertilization, the resulting zygotes (bottom right, day 49) were cultured in M16 medium. All cultures were incubated at 37 °C in an atmosphere of 5% CO₂ in air. Scale bars: white, 500 μm ; black, 50 μm .

same stage (see supplementary information), indicating that normal imprinting can be established *in vitro*.

We also investigated whether nuclear reprogramming could occur in these oocytes by monitoring their development (see supplementary information). Because oocytes isolated from cultured follicles were unable to resume meiosis (0/38, 0%), we transferred the nuclei into enucleated, fully grown oocytes from adult mice. These reconstituted oocytes were able to resume meiosis and mature into metaphase in the second meiosis (M II, 101/108, 94%) and had a normal karyotype ($n = 20$, 11/11, 100%), indicating that chromosomal maturation for meiosis can be completed *in vitro*. As expected⁴, oocytes reconstituted with intact female germ cells from fetuses at 12.5 d.p.c. became aneuploid (32/32, 100%) owing to abnormal chromosomal segregation.

As oocytes reconstituted by a single nuclear transfer did not develop efficiently into blastocysts (14/34, 41%), we used serial nuclear transfer (see supplementary information for details). The rate of *in vitro* fertilization of these reconstituted oocytes was normal (72/81, 89%) and they developed efficiently into blastocysts (64/72, 89%).

After embryo transfer to a surrogate mother, 16 living pups were obtained from 7 mothers at 19.0 d.p.c. by caesarean section (16/64, 25%). No obvious abnormality was seen in any of the pups (*in vitro* pup weight compared with that *in vivo*: 1.46 g versus 1.34 g, $P > 0.1$) or placentae (*in vitro* weight compared with that *in vivo*: 133 mg versus 126 mg, $P > 0.5$), as expected from normal imprinted methylation of *Igf2r*, *Snrpn* and *Peg1* (see supplementary information)^{9,10}. These animals were fertile after puberty.

We have shown that the most primitive murine fetal oocytes can differentiate into competent oocytes with high efficiency. As well as offering an opportunity to analyse the mechanisms behind nuclear reprogramming *in vitro*, our system might eventually help women undergoing chemotherapy or radiotherapy to become mothers afterwards, by prior removal of an ovary.

Yayoi Obata*, Tomohiro Kono†, Izuho Hatada*

*Gene Research Center, Gunma University, 3-39-22 Showa-machi, Maebashi, Gunma 371-8511, Japan
e-mail: ihatada@showa.gunma-u.ac.jp

†Department of Bioscience, Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo 156-8502, Japan

1. Buehr, M. & McLaren, A. *Gamete Res.* **11**, 271–281 (1985).
2. Eppig, J. J. & O'Brien, M. J. *Biol. Reprod.* **54**, 197–207 (1996).
3. Cortvriendt, R., Smits, J. & Van Steirteghem, A. C. *Hum. Reprod.* **11**, 2656–2666 (1996).
4. Bao, S., Obata, Y., Carroll, J., Domeki, I. & Kono, T. *Biol. Reprod.* **62**, 616–621 (2000).
5. Kono, T., Obata, Y., Yoshimizu, T., Nakahara, T. & Carroll, J. *Nature Genet.* **13**, 91–94 (1996).
6. Obata, Y. *et al.* *Development* **125**, 1553–1560 (1998).
7. Obata, Y. & Kono, T. *J. Biol. Chem.* **277**, 5285–5289 (2002).

8. Wutz, A. *et al.* *Nature* **389**, 745–749 (1997).
9. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. *Science* **294**, 2536–2539 (2001).
10. Lucifero, D., Mertineit, C., Clarke, H. J., Bestor, T. H. & Trasler, J. M. *Genomics* **79**, 530–538 (2002).
11. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. *FEBS Lett.* **407**, 313–319 (1997).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Gene silencing

Trans-histone regulatory pathway in chromatin

The fundamental unit of eukaryotic chromatin, the nucleosome, consists of genomic DNA wrapped around the conserved histone proteins H3, H2B, H2A and H4, all of which are variously modified at their amino- and carboxy-terminal tails to influence the dynamics of chromatin structure and function^{1,2} — for example,

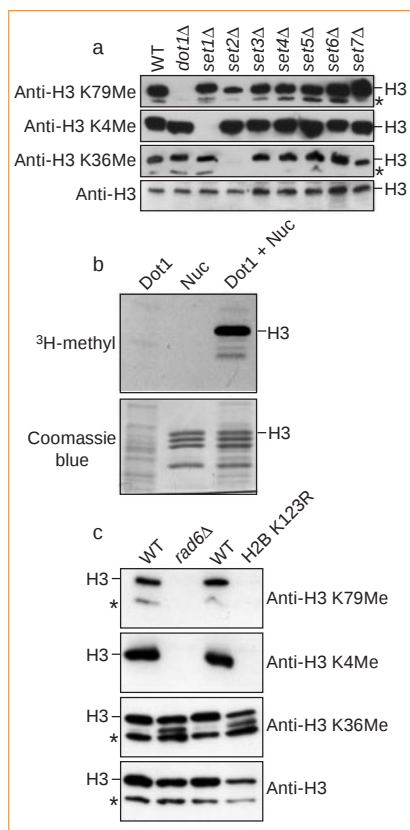


Figure 1 Regulation of Lys-79 methylation on histone H3. **a**, Western blot of nuclear extracts isolated from mutant yeast strains carrying the indicated deletions and probed with antibodies against the methylated H3 lysine residues K79Me, K4Me or K36Me. **b**, Polyacrylamide-gel electrophoresis analysis of recombinant Dot1 for histone methyltransferase activity on nucleosomal substrates (Nuc) *in vitro*, as revealed by autoradiography (top) and Coomassie-blue staining of histones (bottom). **c**, Western blot of nuclear extracts from wild-type yeast and *rad6*-deleted or H2B K123R mutant strains probed with the histone antibodies indicated. Asterisks, an H3 proteolysis product that removes the methylation site at Lys 4 but not at Lys 36 or Lys 79.

conjugation of histone H2B with ubiquitin controls the outcome of methylation at a specific lysine residue (Lys 4) on histone H3, which regulates gene silencing in the yeast *Saccharomyces cerevisiae*³. Here we show that ubiquitination of H2B is also necessary for the methylation of Lys 79 in H3, the only modification known to occur away from the histone tails, but that not all methylated lysines in H3 are regulated by this 'trans-histone' pathway because the methylation of Lys 36 in H3 is unaffected. Given that gene silencing is regulated by the methylation of Lys 4 and Lys 79 in histone H3, we suggest that H2B ubiquitination acts as a master switch that controls the site-selective histone methylation patterns responsible for this silencing.

Lysine residues subject to methylation in yeast histone H3 are Lys 4 and Lys 36 near the amino terminus, and Lys 79 (refs 4–6, and data not shown), a modification site that is unique in that it is located away from the H3 tails in the first loop of the histone-fold domain. To identify what mediates methylation of Lys 79, we used an antibody raised against this methylated site (anti-H3 K79Me; Upstate Biotechnology) to screen nuclear extracts isolated from yeast strains containing deletions of known and putative histone methyltransferase enzymes.

We identified Dot1, a factor involved in gene silencing^{4,5}, as a gene product that is essential for methylation of Lys 79 (Fig. 1a), in agreement with earlier findings^{4–6}. This was unexpected, as Dot1 lacks the 'SET' domain, which until now was thought to be the only domain responsible for methylating histone lysine residues. No other protein containing the SET domain was found to mediate Lys-79 methylation (Fig. 1a). To confirm that Dot1 is the enzyme responsible, we showed that expression of *DOT1* in a *dot1*-deleted strain restores Lys-79 methylation in H3 (see supplementary information) and that recombinant Dot1 contained methyltransferase activity towards nucleosomal H3 (Fig. 1b). Dot1, which resembles the arginine methyltransferase family in sequence and structure⁷, therefore represents a new class of lysine-specific histone methyltransferase enzymes.

Ubiquitination of histone H2B is mediated by the enzyme Rad6, also known as ubiquitin-conjugating enzyme Ubc2 (ref. 8). We tested whether Rad6 ubiquitination of H2B could influence methylation at H3

sites apart from Lys 4 (ref. 3). Surprisingly, we found that there was a loss of Lys-79 methylation, but not of Lys-36 methylation, in strains deleted for *RAD6* or mutated at the H2B-ubiquitination site (K123R; Fig. 1c). We verified that Dot1 is expressed in these mutant strains by using reverse transcription followed by polymerase chain reaction to detect the presence of its messenger RNA (see supplementary information). We conclude that Rad6 ubiquitination of H2B at Lys 123 specifically regulates the methylation of both Lys 4 and Lys 79 of H3 in a 'trans-histone' pathway.

We found that deletion of *DOT1*, *SET1* (refs 9, 10) or *SET2* (ref. 11) results in the specific loss of their respective modifications (Fig. 1a), indicating that the absence of these individual modifications does not affect the others. Given that the regulation of Lys-4 methylation by H2B ubiquitination is unidirectional³ and that Lys-4 methylation still occurs in the *dot1*-deleted (Fig. 1a) and Lys-79 mutant strains (see supplementary information), we conclude that the regulation of Lys-79 methylation by H2B ubiquitination is also unidirectional.

Our findings indicate that methylation of histone H3 at Lys 4 and Lys 79, but not at Lys 36, is regulated by the ubiquitination of histone H2B. As Lys 4 and Lys 79 methylation both mediate gene silencing, we propose that H2B ubiquitination acts as a master switch of gene silencing through a *trans*-histone pathway that leads to the appropriate patterns of histone methylation. Given that some lysine residues in H3 are affected by this pathway, but not others, our results lend support to the 'histone code' hypothesis².

Scott D. Briggs*, **Tiaojiang Xiao†**, **Zu-Wen Sun***, **Jennifer A. Caldwell‡**, **Jeffrey Shabanowitz‡**, **Donald F. Hunt‡**, **C. David Allis***, **Brian D. Strahl†**

*Department of Biochemistry and Molecular Genetics, University of Virginia Health System, Charlottesville, Virginia 22908, USA

†Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA
e-mail: brian_strahl@med.unc.edu

‡Departments of Chemistry and Pathology, University of Virginia, Charlottesville, Virginia 22908, USA

1. Kornberg, R. D. & Lorch, Y. *Cell* **98**, 285–294 (1999).
2. Strahl, B. D. & Allis, C. D. *Nature* **403**, 41–45 (2000).
3. Sun, Z. W. & Allis, C. D. *Nature* **418**, 104–108 (2002).
4. van Leeuwen, F., Galken, P. R. & Gottschling, D. E. *Cell* **109**, 745–756 (2002).
5. Ng, H. H. *et al.* *Genes Dev.* **16**, 1518–1527 (2002).
6. Feng, Q. *et al.* *Curr. Biol.* **12**, 1052–1058 (2002).
7. Dlakic, M. *Trends Biochem. Sci.* **26**, 405–407 (2001).
8. Robzyk, K., Recht, J. & Osley, M. A. *Science* **287**, 501–504 (2000).
9. Briggs, S. D. *et al.* *Genes Dev.* **15**, 3286–3295 (2001).
10. Bryk, M. *et al.* *Curr. Biol.* **12**, 165–170 (2002).
11. Strahl, B. D. *et al.* *Mol. Cell Biol.* **22**, 1298–1306 (2002).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Published online 14 July 2002; doi:10.1038/nature00970

Supernova-remnant origin of cosmic rays?

It is thought that Galactic cosmic-ray nuclei are gradually accelerated to high energies (up to about 300 TeV per nucleon, where 1 TeV is 10^{12} eV) in the expanding shock waves connected with the remnants of powerful supernova explosions. However, this conjecture has eluded direct observational confirmation^{1,2} since it was first proposed in 1953 (ref. 3). Enomoto *et al.*⁴ claim to have finally found definitive evidence that corroborates this model, proposing that very-high-energy, TeV-range, γ -rays from the supernova remnant (SNR) RX J1713.7–3946 are due to the interactions of energetic nuclei in this region. Here we argue that their claim is not supported by the existing multiwavelength spectrum of this source. The search for the origin(s) of Galactic cosmic-ray nuclei may be closing in on the long-suspected supernova-remnant sources, but it is not yet over.

We have previously suggested⁵ that the SNR RX J1713.7–3946 (dark X-ray contours in Fig. 1) might be an accelerator of cosmic-ray nuclei on the basis of GeV (10^9 eV) γ -ray emission⁶ (Fig. 1, white contours) seen towards adjacent massive and dense molecular clouds (Fig. 1, false colour), with which it seems to be interacting⁷; the γ -ray signature of nucleonic interactions is known to be greatly amplified in such dense media. Enomoto *et al.* have since claimed⁴ that the γ -rays of even higher energy (TeV range) that they detected directly towards

the more rarefied remnant (red contours in Fig. 1) are an unambiguous signature of nucleonic interactions in this region, thus finally proving the long-held conjecture.

However, it is well known that SNRs accelerate electrons (see ref. 8, for example) that can also generate TeV γ -rays at the source sites. It is therefore generally difficult to assign conclusively an electronic rather than a nucleonic origin to the observed celestial γ -rays⁹. Fortunately, in this case the discrimination is not so subtle — if the TeV emission seen in the northwest quadrant of the remnant were really due to protons, as proposed⁴, that part of the SNR would have been detected as a bright GeV source in the Energetic Gamma Ray Experiment Telescope (EGRET) all-sky survey⁶, but it was not.

The proposed model⁴ is unsatisfactory because it predicts that there will be roughly three times the GeV γ -ray intensity of a nearby EGRET source⁶ (3EG 1714–3857) at a location where no actual GeV flux has been detected (that is, above the background level), although alternative models might reduce the discrepancy (see also ref. 10). In addition, the high gas density (about 10^2 protons cm^{-3}) needed to explain the TeV γ -rays as nucleonic in origin⁴ is simply not compatible — by several orders of magnitude — with the X-ray data⁷ from the same region of the SNR.

As the peaks of both the TeV (ref. 4) and X-ray (ref. 7) emissions lie in the same low-density⁷ region of the remnant (Fig. 1), the most plausible explanation is that they both arise from the same population of relativistic electrons. However, there is an exciting possibility that a minor component of the TeV emission⁴ could be due to nucleonic

interactions at the location of the adjacent cloud B (Fig. 1), where sufficient molecular material is present⁵. A suggestive extension and a local maximum in the TeV significance contours⁴ are coincident with this location.

SNR RX J1713.7–3946 might therefore be accelerating nuclei to cosmic-ray energies⁵, but the TeV γ -ray signature so far detected from the northwest quadrant of the remnant⁴ is not proof of this. It is likely that most — but not all — of the observed TeV emission is actually electronic in origin.

To address the question of the source(s) of the TeV emission properly, better radio data sampled at several frequencies, as well as the inclusion of existing low-energy ROSAT X-ray data, will be needed to determine the primary synchrotron spectrum. Data of greater spatial resolution from the next-generation GeV, TeV and neutrino telescopes will also be important. Only when this multifrequency data set is available can the nature of the contributions (nucleonic or electronic) to the observed TeV emissions start to be disentangled.

Yousaf M. Butt*, Diego F. Torres†, Gustavo E. Romero‡, Thomas M. Dame*, Jorge A. Combi‡

*Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA
e-mail: ybutt@cfa.harvard.edu

†Department of Physics, Jadwin Hall, Princeton University, Princeton, New Jersey 08544, USA

‡Instituto Argentino de Radioastronomía, CC 5, 1894, Villa Elisa, Buenos Aires, Argentina

1. Drury, L. O'C. *et al.* *Space Sci. Rev.* **99**, 329–352 (2001).
2. Plaga, R. Preprint astro-ph/0111555 at <http://xxx.lanl.gov> (2001).
3. Shklovskii, I. S. *Dokl. Akad. Nauk SSSR* **91**, 475–478 (1953).
4. Enomoto, R. *et al.* (CANGAROO collaboration) *Nature* **416**, 823–826 (2002).
5. Butt, Y. M., Torres, D. F., Combi, J. A., Dame, T. & Romero, G. E. *Astrophys. J.* **562**, L167–L171 (2001).
6. Hartman, R. C. *et al.* *Astrophys. J.* (suppl.) **123**, 79–202 (1999).
7. Slane, P. *et al.* *Astrophys. J.* **525**, 357–367 (1999).
8. Koyama, K. *et al.* *Nature* **378**, 255–258 (1995).
9. Pohl, M. *Astron. Astrophys.* **307**, L57–L59 (1996).
10. Reimer, O. & Pohl, M. Preprint astro-ph/0205256 at <http://xxx.lanl.gov> (2002); *Astron. Astrophys.* (in the press).

Does similarity breed cooperation?

Reciprocity¹, whether direct² or indirect³, is thought to be the key to establishing cooperation among non-relatives. But Riolo *et al.*⁴ have presented a model in which cooperation is instead based on similarity: agents donate only when their partner's 'tag' lies within a 'tolerance' range around their own. Here we point out that their model requires individuals with identical tags to cooperate with each other, and show that cooperation tends to collapse when individuals bearing identical tags are given the

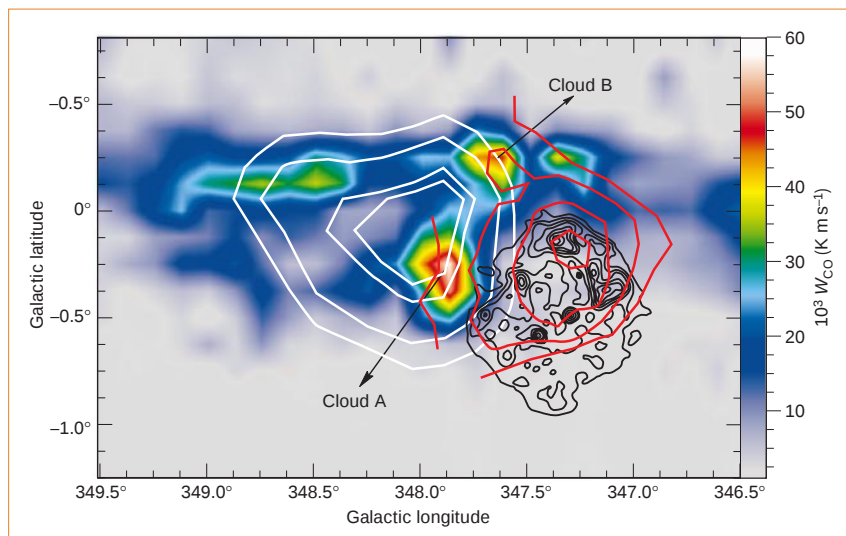


Figure 1 An overlay map in Galactic coordinates showing supernova remnant (SNR) RX J1713.7–3946 (G347.3–0.5) in dark contours (ROSAT PSPC X-ray; from ref. 7). Red, TeV significance contours⁴; white, location-probability contours of the GeV (10^9 eV) EGRET source 3EG J1714–3857 (ref. 6); colour scale indicates the intensity of CO ($J=1 \rightarrow 0$) emission, and consequently the column density of the ambient molecular clouds, in the velocity interval $v = -105$ to -80 km s^{-1} associated with the SNR, corresponding to a kinematic distance of 6.3 ± 0.4 kpc (ref. 5). Note the local maximum of the TeV significance contours at cloud B, which may indicate a hadronic origin of the TeV flux at that location. For further details, see ref. 5.

option of not donating. We therefore question their mechanism for maintaining cooperation without reciprocity.

What makes cooperation so challenging for theorists is explaining how it can persist in the face of more exploitative strategies. In the system of Riolo *et al.*, clusters of cooperating agents with similar tags arise intermittently, only to be undermined by agents that reduce their tolerance level, T , such that they accept more donations than they offer. However, there is a limit to such cheating imposed by the minimum $T = 0$. This means that when individuals with identical tags interact, they must always donate. A striking characteristic of Riolo and colleagues' simulations was the formation, through differential reproduction, of clusters of agents with identical tags. Most individuals in their simulations (up to 97% of the population⁴) were therefore ultimately constrained to cooperate.

To investigate what would happen if agents were given the option of declining to donate to any other agents, even those with identical tags, we replicated Riolo and colleagues' simulations with one simple modification: we allowed tolerance to evolve to below zero. Agents with negative T values would not donate to any other agent, although, by setting the minimum boundary for T at -10^{-6} , we ensured that all positive mutational changes converted T back to the cooperative region. We found that introducing the realistic option of non-donation had a catastrophic effect on cooperation (Fig. 1).

Why do we not find the high degree of cooperation reported by Riolo *et al.*? Once the constraint that identical tags must cooperate has been removed, agents interacting with others bearing the same tag face the classical 'prisoner's dilemma' — they can do well by cooperating, but they can do even better by accepting donations without donating. Thus, mutants that fail to donate, even to those with identical tags, will tend to invade, destroying cooperation.

Cooperation under the original conditions of Riolo *et al.* operates through a process of 'like helping like'⁵ — agents sharing any particular tag also share the rule of donating to each other, so a form of kin selection⁶ can support cooperation. However, agents can have identical tags without having a recent common ancestor, so in our modified system they can share tags without sharing the rule for cooperating. Because tag similarity is no longer a reliable guide to behaviour, the system of 'like helping like' breaks down. Whereas the problematic 'green beard' effect⁷ depends on a link between altruism and a particular trait, the system of Riolo *et al.* depends on a link between altruism and similarity. Allowing similar individuals not to donate caused cooperation to be restricted in our system, even without any mechanism for cheating through faking tags⁵.

Cooperation based on similarity there-

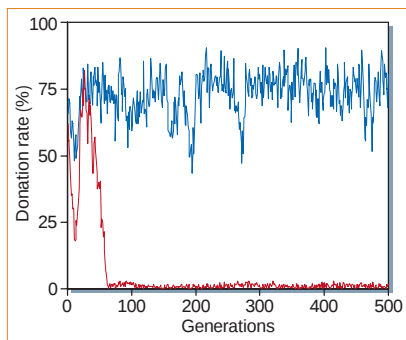


Figure 1 Population dynamics for the first 500 generations of a typical run of Riolo and colleagues' model⁴, in which individuals with identical tags must donate (blue), and our modified model in which individuals with identical tags may or may not donate (red). All parameter values were the same as for Fig. 1 of ref. 4. In 30 runs of our modified model, each for 30,000 generations, the overall mean donation rate was 1.48% (s.e.m. 0.031%), in comparison with Riolo *et al.*'s 73.6%.

fore turns out to be a rule that was built into the model rather than an inference that can be drawn from it. Nevertheless, we believe that possible mechanisms by which cooperation can arise without reciprocity merit further attention⁸, and the role of signals in such systems will be an important consideration.

Gilbert Roberts*, Thomas N. Sherratt†‡

*Evolution and Behaviour Research Group, School of Biological Sciences, Henry Wellcome Building for Neuroecology, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK

e-mail: gilbert.roberts@ncl.ac.uk

†School of Biological and Biomedical Sciences, University of Durham, South Road, Durham DH1 3LE, UK

‡Present address: Department of Biology, Carleton University, Ottawa, Ontario K1S 5B6, Canada

1. Trivers, R. Q. *Rev. Biol.* **46**, 35–57 (1971).
2. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390–1396 (1981).
3. Nowak, M. A. & Sigmund, K. *Nature* **393**, 573–577 (1998).
4. Riolo, R. L., Cohen, M. D. & Axelrod, R. *Nature* **414**, 441–443 (2001).
5. Sigmund, K. & Nowak, M. A. *Nature* **414**, 403–405 (2001).
6. Hamilton, W. D. *J. Theor. Biol.* **7**, 1–52 (1964).
7. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, New York, 1976).
8. Roberts, G. *Proc. R. Soc. Lond. B* **265**, 427–431 (1998).

Riolo *et al.* reply — Roberts and Sherratt argue that if agents with identical tags are allowed a choice of behaviour, then tag similarity can no longer be a reliable guide to behaviour and so similarity does not breed cooperation. Although they are correct in noting that in our model¹ an agent will always donate when it meets another with an identical tag, we do not believe that their basic claim is correct.

We have replicated the results of Roberts and Sherratt and have run a generalized model that includes theirs as one extreme and our original model as another (details are available from R.L.R.). We find that if mutations are not biased as strongly

towards 'never donate', as in their version of the model, similarity can indeed breed cooperation. Whether it does, and to what extent, depends on several factors, including the rate at which 'never donate' agents are created, the number of pairings, the cost/benefit ratio of donations and the particular adaptive mechanisms in the model. If unconditional defection is introduced by adding a binary trait that controls whether agents never donate, or donate using tags and tolerance, we find that cooperation also emerges, but again the extent of cooperation depends on many factors.

We believe that the difference has not been fully understood between the stability of cooperation within any particular tag group and the rate of cooperation across a population consisting of diverse tags with changing frequencies over time. There is no dispute that particular cooperative tag groups are invadable^{1,2}. However, as one tag group is invaded and thus dies off, another tag group with more reliable cooperators can flourish and become dominant, resulting in the cycles of cooperation and tag dominance noted previously^{1,3}.

Roberts and Sherratt claim that cooperation based on similarity was built into our model. It was not, which is why, under some parameter settings (few pairings or high cost of donation), cooperative periods are rare and short-lived, resulting in very low overall donation rates¹. Nevertheless, the level of cooperation for other parameter settings is substantial, with the overall rate of cooperation depending on the relative dynamics of invasion, resistance and emergence of dominant groups.

Many factors could affect the dynamics generated by tag-based mechanisms. For instance, tags that are easy to copy might lead to high rates of invasion, whereas other tags, such as language or accent, might be difficult to copy³. Our model could also be extended to study how a tag mechanism acts in conjunction with other mechanisms known to affect the emergence of cooperation. For example, territorial distribution of agents might favour 'speciation' into self-enforcing stereotypes³. Further investigation is needed to understand fully the range of mechanisms that can produce cooperation without reciprocity. Our results show that, under some conditions, tag mechanisms are one viable approach.

Rick L. Riolo*, Michael D. Cohen†, Robert Axelrod‡

*Center for the Study of Complex Systems, †School of Information, and ‡Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor, Michigan 48109, USA
e-mail: rlrilo@umich.edu

1. Riolo, R. L., Cohen, M. D. & Axelrod, R. *Nature* **414**, 441–443 (2001).
2. Robson, A. J. *J. Theor. Biol.* **144**, 379–396 (1990).
3. Sigmund, K. & Nowak, M. A. *Nature* **414**, 403–405 (2001).

Functions of FGF signalling from the apical ectodermal ridge in limb development

Xin Sun*, Francesca V. Mariani & Gail R. Martin

Department of Anatomy and Program in Developmental Biology, School of Medicine, University of California at San Francisco, San Francisco, California 94143-0452, USA

To determine the role of fibroblast growth factor (FGF) signalling from the apical ectodermal ridge (AER), we inactivated *Fgf4* and *Fgf8* in AER cells or their precursors at different stages of mouse limb development. We show that FGF4 and FGF8 regulate cell number in the nascent limb bud and are required for survival of cells located far from the AER. On the basis of the skeletal phenotypes observed, we conclude that these functions are essential to ensure that sufficient progenitor cells are available to form the normal complement of skeletal elements, and perhaps other limb tissues. In the complete absence of both FGF4 and FGF8 activities, limb development fails. We present a model to explain how the mutant phenotypes arise from FGF-mediated effects on limb bud size and cell survival.

A fundamental question in developmental biology is how structures of various sizes and shapes are formed. The vertebrate limb has long been considered an excellent model system for addressing this question. The challenge has been to understand how a simple embryonic bud, containing morphologically homogeneous mesenchymal cells and a covering epithelium, develops into an organ that contains numerous elements with diverse forms^{1,2}. Experiments first performed more than 50 years ago³ identified the ectoderm rimming the distal tip of the limb bud (apical ectodermal ridge, AER) as a source of signals essential for limb development. When the AER is removed at an early limb bud stage, the most proximal skeletal segment (stylopod: upper arm, thigh) forms, but middle (zeugopod: forearm, lower leg) and distal (autopod: wrist and hand, ankle and foot) segments are absent. When it is removed at a slightly later stage, only the autopod is missing^{3–5}. Swapping AERs of early and late stage limb buds has no effect on limb skeletal pattern⁶, establishing that AER signals are permissive but not instructive for limb development.

Experiments showing that distal skeletal elements could be rescued by applying beads soaked in recombinant FGF protein (FGF beads) to the tip of AER-denuded chick limb bud suggested that members of the FGF family of secreted proteins are the AER-derived signals required for limb development^{7,8}. Four of the 22 known FGF genes⁹, *Fgf4*, *Fgf8*, *Fgf9* and *Fgf17*, display AER-specific expression domains within the mouse limb bud (AER-FGFs)^{10,11}. Genetic analysis of *Fgf4* and *Fgf8* function in the limb requires a tissue-specific inactivation approach, because null homozygosity for either gene causes embryonic lethality before limbs develop^{12,13}. In contrast, embryos homozygous for null alleles of either *Fgf9* or *Fgf17* are viable^{14,15}. Individual loss of function of *Fgf4* (refs 11, 16), *Fgf9* (ref. 14), or *Fgf17* (ref. 15) has no effect on limb development, whereas inactivation of *Fgf8* can cause hypoplasia of all three limb segments^{17,18}. Here, we found that inactivating both *Fgf4* and *Fgf8* in the AER produced more severe phenotypes than inactivating *Fgf8* alone. The defects we observed can be explained by the hypothesis that a principal function of AER-FGF signalling during normal limb development is to ensure that enough progenitor cells are available to form each element of the limb skeleton¹⁷, and perhaps other tissues as well.

Timing of gene inactivation differs in hindlimbs and forelimbs

In the animals that we studied (Fig. 1a), hereafter described as *Msx2-cre; Fgf4; Fgf8* or double knockout mice, conditional (flanked by *loxP* sites, floxed) alleles of *Fgf4* (ref. 11) and *Fgf8* (ref. 19) are converted to null alleles by Cre protein produced by the *Msx2-cre* transgene. Notably, *Msx2-cre* is expressed differently in hindlimb compared with forelimb ectoderm¹¹. In nascent hindlimb buds of *Msx2-cre; Fgf4* and *Msx2-cre; Fgf8* single knockout embryos, *Msx2-cre* functions early enough to cause complete inactivation of *Fgf4* and *Fgf8* before they are ever expressed^{11,17}. Thus, we detected neither *Fgf4* nor *Fgf8* RNA in nascent double knockout hindlimb buds using probes that hybridize to sequences deleted by Cre, even at somite stages (som) when *Fgf4* and *Fgf8* expression is normally commencing in hindlimb buds (34 somites, about embryonic day (E)10.5, and 27 som, about E9.5, respectively) (Fig. 1b–e, n). Furthermore, we detected no FGF8 protein in double knockout hindlimb buds at 30 som or later stages (Fig. 1f, g, and data not shown). A similar analysis of FGF4 protein could not be performed because a suitable antibody was not available. These results indicate that double knockout hindlimb buds form in the complete absence of FGF4 and FGF8 activity.

In contrast, in forelimb buds, *Msx2-cre* functions after expression of *Fgf8* has been initiated. Notably, in *Fgf8* single knockout forelimb buds, transient *Fgf8* expression followed by inactivation causes abnormally early expression of *Fgf4* (ref. 17). Consistent with these observations, in double knockout forelimb buds at 27 som we detected *Fgf8* expression, but at markedly reduced levels, and precocious expression of *Fgf4* (Fig. 1h–k). However, by 29 som and at later stages, neither *Fgf4* nor *Fgf8* RNA was detected (not shown). These data demonstrate that both *Fgf4* and *Fgf8* are inactivated in double knockout forelimb buds, but only after expression of both genes has commenced, and thus early forelimb development proceeds in the presence of some FGF4 and FGF8 activity (Fig. 1n). This activity, however, appears to be short lived, as FGF8 protein was not detected in the double knockout forelimb AER at 30 som, shortly after *Fgf8* RNA became undetectable (Fig. 1j–m, and data not shown).

Skeletal phenotypes differ in hindlimbs and forelimbs

None of 23 newborn double knockout mice examined had hind-

* Present address: Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA.

limbs. Where hindlimbs should have formed, small cartilaginous rudiments were observed (Fig. 2a, b), except in four cases that had no skeletal tissue (not shown). In contrast, forelimbs of double knockout mice contained elements of all three limb segments. Stylopod (humerus) was invariably present, but was sometimes shorter or thinner than normal (14 out of 46 forelimbs) (Fig. 2c–e). Zeugopod elements (radius, ulna) were always severely hypoplastic and malformed. Often, only a single element was present (25 out of 46), which usually could be identified as ulna (Fig. 2d, e). Autopod (wrist, hand plate) appeared to lack many wrist bones, and those present were fused (Fig. 2f–h). Only one (24 out of 46), two (20 out of 46), or at most three digits (2 out of 46) were present (Fig. 2g, h, and data not shown). Digits usually contained two phalanges, and thus represented either digit 1 or other digits with a missing phalanx. Terminal phalanges displayed normal distal morphology (Fig. 2f–h). These data show that when *Fgf4* and *Fgf8* are never expressed in limb bud ectoderm, there is virtually complete failure of skeletal development; however, when they are transiently active at very early limb bud stages, the most proximal segment develops

normally but both the middle and distal segments are severely hypoplastic.

To determine whether these defects were due to reduction in size of the precartilagenous condensations that prefigure the skeletal elements, we assayed for *Sox9* RNA²⁰ (Fig. 2i–p). As early as 38 som, *Sox9*-expressing cell populations in double knockout limb buds were much smaller than normal (Fig. 2i, j, m, n). By 46 som, when *Sox9*-positive condensations representing each segment can be distinguished, precursor populations of all three segments were absent or substantially reduced in size in double knockout hindlimb buds (Fig. 2k, l). In contrast, in forelimb buds, the stylopod precursor population appeared normal in size, whereas zeugopod and autopod precursor populations were much smaller than normal (Fig. 2o, p). These results suggest that the double knockout limb phenotype is due to a decrease in skeletal progenitor cell number, at or before the condensation stage. Figure 2k, l, o, p also illustrates the extent to which double knockout limb buds are reduced in overall size by approximately E11.75 (46 som).

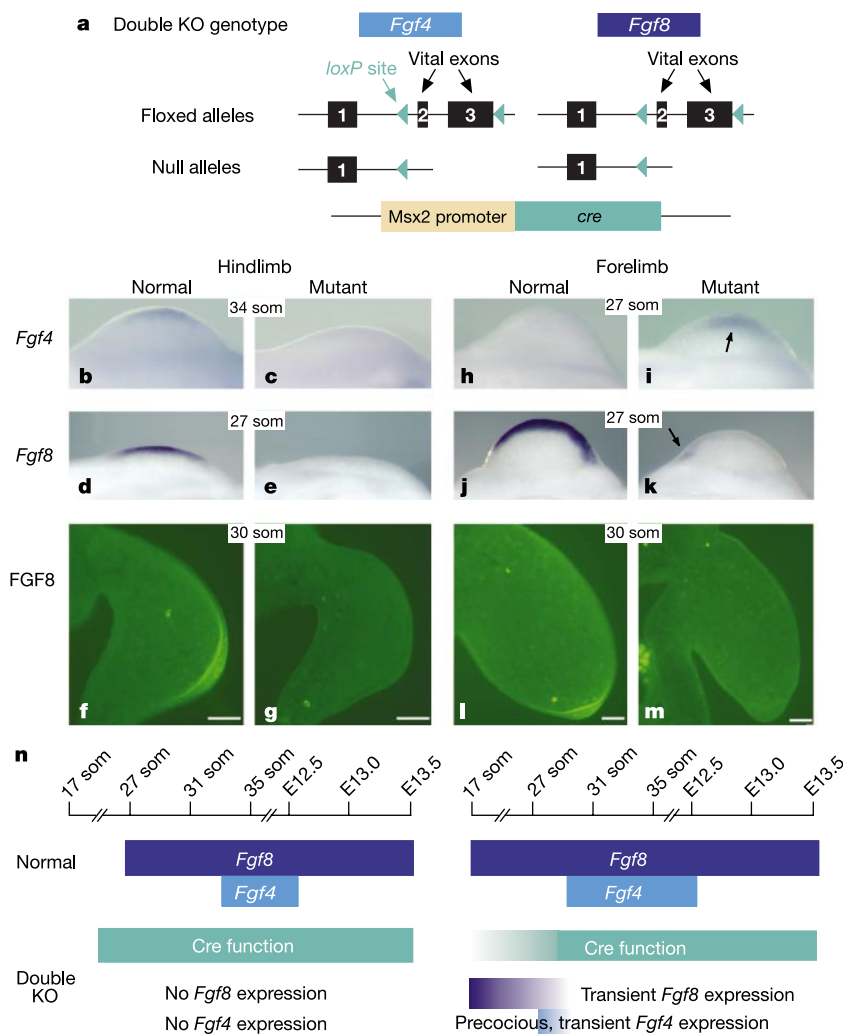


Figure 1 Timing of *Fgf4* and *Fgf8* inactivation in hindlimb and forelimb buds. **a**, Diagram of mutant alleles carried by *Msx2-cre; Fgf4; Fgf8* (double knockout, KO) animals. Cre recombinase, expressed in limb bud ectoderm under the control of *Msx2* promoter elements, mediates intragenic recombination of *loxP* sites flanking *Fgf4* and *Fgf8* sequences. This deletes essential exons in the floxed alleles, converting them to null alleles. **b–m**, Analysis of gene expression and protein levels in hindlimb (**b–g**) and forelimb (**h–m**) buds. Limb bud developmental age is reported in terms of somite number (som). Intact limb buds (**b–e**, **h–k**) are shown in dorsal view, with anterior on the left and

posterior on the right. Transverse sections (**f**, **g**, **l**, **m**) are shown with dorsal at the top, proximal to the left. Relative sizes of the limb buds are represented by white scale bars. Assays were performed for *Fgf4* (**b**, **c**, **h**, **i**) and *Fgf8* (**d**, **e**, **j**, **k**) expression by RNA *in situ* hybridization in whole mount, and for FGF8 protein (**f**, **g**, **l**, **m**) by immunofluorescence. The arrows in **i** and **k** point to precocious *Fgf4* expression and the abnormally small domain of *Fgf8* expression, respectively. **n**, Diagram illustrating the timing of *Fgf4* and *Fgf8* expression and the timing of *Msx2-cre* function.

AER development and mesenchymal gene expression

A lack of skeletal development occurs when the AER fails to form or degenerates at a very early stage^{21–24}. This was not, however, the cause of the hindlimb aplasia that we observed, as double knockout hindlimb buds contained a morphologically normal AER (Fig. 3a, b, and data not shown). It remained intact until 42 som (about E11.25), when portions began to degenerate as a result of higher than normal levels of cell death (see below). At approximately E10.5 (34–36 som), we observed relatively normal expression of AER genes, including *Fgf9*, *Jag2*, *Bmp4*, *Bmp2*, *Msx2* and even *Fgf8* itself, as detected with a probe for *Fgf8* exon 1 sequences not deleted by Cre. *Fgf17* expression was also detected, but at a slightly higher level than normal (Fig. 3c–j, and data not shown). Assays at earlier stages showed that neither *Fgf9* nor *Fgf17* was precociously expressed (not shown). In double knockout forelimb buds (see Supplementary Information), the AER was initially morphologically normal, but subsequently developed patches of abnormal thickness. AER gene expression reflected this patchy hyperplasia, but was otherwise normal until about E11.25, when the AER began to degenerate. Thus, our data indicate that a relatively normal AER can be

established and is maintained until late limb bud stages in the absence of FGF4 and FGF8 activities.

In double knockout hindlimb buds, mesenchyme gene expression was severely affected by the lack of *Fgf4* and *Fgf8* function. For example, *Bmp2*, *Bmp4* and *Msx2* RNAs could not be detected in the mesenchyme, even though they were detected in their normal domains in the AER (Fig. 3i, j, and data not shown). Similarly, expression of *Hoxd13*, *Fgf10*, *Spry2*, *Spry4*, as well as *Shh* and its downstream target gene *Ptch* was not detected in double knockout hindlimb buds at about E10.5 (Fig. 3k–p, and data not shown). Assays at earlier stages (not shown) revealed that *Shh* and *Ptch* expression was never initiated in the double knockout hindlimb buds. Hindlimb mesenchymal cells were capable of gene expression, as gremlin, *Msx1* and *Meis1* RNAs were detected, but in abnormal patterns (Fig. 3q–v). For example, gremlin expression failed to be upregulated, as in normal limb buds (Fig. 3q, r). *Alx4* expression, which normally marks anterior mesenchyme, was detected at higher levels and in a slightly larger domain than normal (Fig. 3w, x), perhaps reflecting the absence of SHH signalling²⁵.

In contrast, in double knockout forelimb buds most mesenchy-

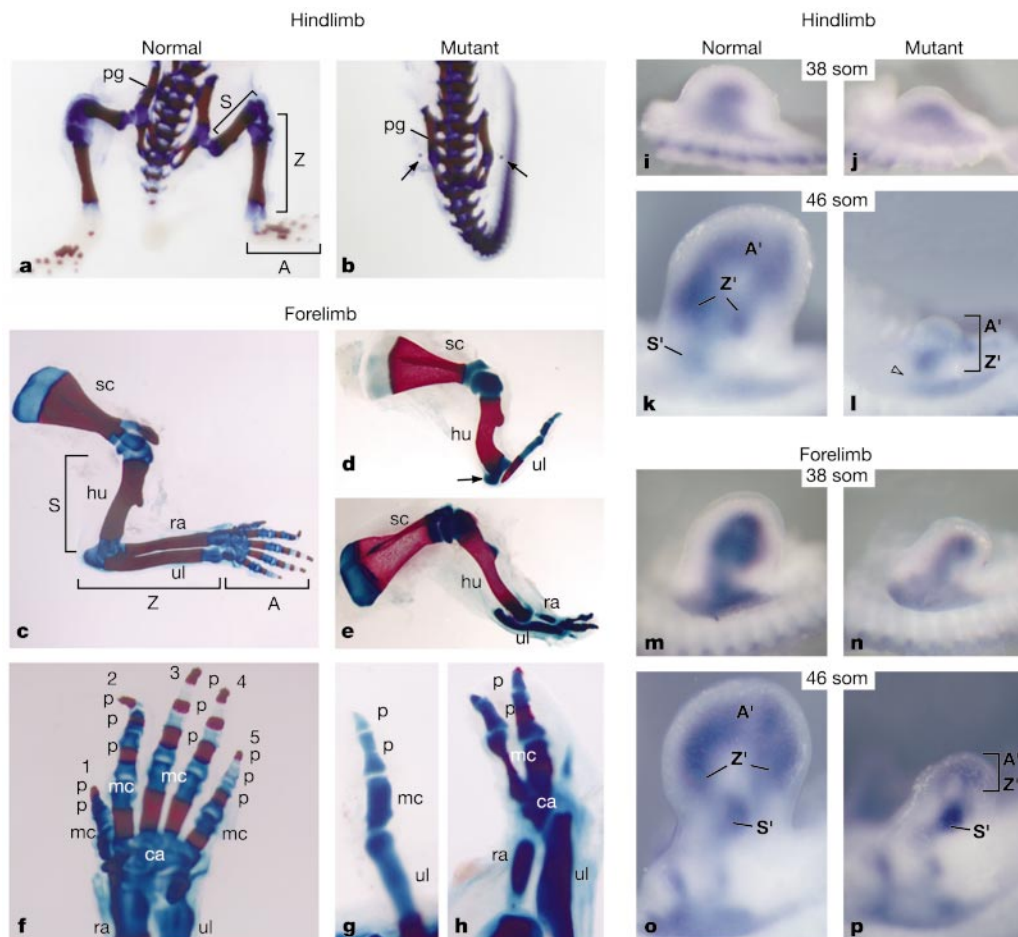


Figure 2 Limb skeletal phenotypes. **a–h**, Skeletal preparations of hindlimbs (**a**, **b**) and forelimbs (**c–h**) from normal and double knockout mice at birth, stained for cartilage (alcian blue) and bone (alizarin red). Arrows in **b** point to vestigial cartilaginous elements present in place of the hindlimb skeleton in double knockout mice. Note that the pelvic girdle, which is not derived from the hindlimb bud, appears normal. Forelimbs of two double knockout mice are shown, illustrating the most severe (**d**) and mildest (**e**) phenotypes obtained. The arrow in **d** points to the olecranon process, which identifies the single zeugopod element as an ulna. Autopods of the skeletons in **c**, **d**, **e**, are shown at higher magnification in **f**, **g**, **h**, respectively. **i–p**, *Sox9* expression assayed by RNA *in situ*

hybridization in whole mount in normal and double knockout limb buds at the stages indicated. The locations of hindlimb skeletal precursors (**k**) was inferred from fate-mapping studies⁴⁶ and histological analyses of wild-type mouse hindlimb buds⁴⁷. The open arrowhead in **l** points to the region in which the stylopod precursor should be located. Note the absence of *Sox9* RNA in this region. A, autopod; A', prospective autopod; ca, carpal; hu, humerus; mc, metacarpal; p, phalanx; pg, pelvic girdle; ra, radius; S, stylopod; S', prospective stylopod; sc, scapula; ul, ulna; Z, zeugopod; Z', prospective zeugopod. Digits in **f** are numbered 1–5 from anterior to posterior, respectively.

mal genes assayed were expressed, but at lower levels than normal, generally in domains that were reduced in size approximately in proportion with the size reduction of the forelimb buds. However, *Alx4* expression appeared to be stronger than normal and expanded posteriorly (see Supplementary Information). Together, these data demonstrate that the combined activities of FGF4 and FGF8 are essential to induce and/or maintain expression of many genes in limb bud mesenchyme. The comparative normality of most gene expression patterns in forelimb bud mesenchyme is presumably due, directly or indirectly, to transient expression of *Fgf4* and *Fgf8* at early stages of development.

Abnormal cell death and normal cell proliferation

A notable feature of double knockout limb buds is their markedly reduced size by about E10.5 (Fig. 3). This might be due to cell death, as AER removal from the chick limb bud results in rapid death of the underlying mesenchyme^{26,27}. Similarly, within 7 h of AER removal from mouse limb buds, many distal mesenchymal cells die (Fig. 4a, b). We assayed for cell death in double knockout limb buds at approximately 2–4-h intervals between 21 som and 45 som by Nile blue sulphate staining or TdT-mediated dUTP nick end labelling (TUNEL). In hindlimb buds, increased cell death was detected, but not until 34 som and never in the distal limb bud. Instead, it was observed in proximal mesenchyme on the dorsal side (Fig. 4c–f, and data not shown). Similarly, beginning at 27 som, only a proximal patch of apoptotic cells was detected in double knockout forelimb buds. Interestingly, this domain of cell death included dorsal ectoderm (Fig. 4g, h, and data not shown). By 42 som, and at later stages, no abnormal mesenchymal cell death was detected in forelimb buds, although it persisted in proximal mesenchyme of double knockout hindlimb buds until at least 42 som. The cell death domain was centrally located along the anterior–posterior axis (Fig. 4i, and data not shown). Furthermore, higher than normal numbers of apoptotic cells were detected in the AER of double knockout hindlimb and forelimb buds beginning at 39 som and 32 som,

respectively, and persisting until at least 42 som (Fig. 4g, h, and data not shown).

These results showing no abnormal cell death in double knockout limb buds at early stages but extensive apoptosis at later stages only in proximal mesenchyme, are consistent with what has been observed in *Fgfr2b*-deficient limb buds in which the AER fails to form²³, and also in *Msx2-cre; Fgf8* single knockout limb buds (M. Lewandoski and G.R.M., unpublished data). However, they contrast with the report of apoptosis in anterior distal mesenchyme in *Fgf8*-deficient limb buds¹⁸. The fact that cell death occurs far from the AER is perplexing, because FGFs are thought to act near the cells that produced them. Thus cell death might be due to an indirect effect of loss of FGF4 and FGF8 signalling.

In trying to identify the apoptotic cells, we considered the possibility that they might be migrating myoblasts. However, expression of *Pax3*, a myoblast marker²⁸, appeared normal in double knockout limb buds (Fig. 4j–m, and data not shown). We also saw no obvious abnormalities in limb bud vasculature, as assessed by staining for PECAM, a vascular endothelial cell marker²⁹ (not shown). Thus it is uncertain whether the dying cells are progenitors of chondrocytes or other cell types in the limb. However, it is noteworthy that the cell death domains include regions of normal cell death (Fig. 4g, h), suggesting that loss of AER-FGF signalling may exacerbate a normal process.

We also examined cell proliferation by assaying for phosphorylated histone H3 (pH3), a marker of mitotic cells³⁰. Interestingly, in normal hindlimb buds, pH3-positive cells appeared evenly distributed throughout the mesenchyme at 35 som (Fig. 5a), suggesting that there is no gross local variation in cell proliferation at this stage. Similarly, the mitotic index was the same in distal and proximal chick limb bud mesenchyme, at least until about 45 som (Hamburger–Hamilton stage 23)³¹. In double knockout hindlimb buds, pH3-positive cells were evenly distributed throughout the mesenchyme, at the same density as in normal hindlimb buds (Fig. 5a, b). Using confocal microscopy, we quantified cell proliferation in distal

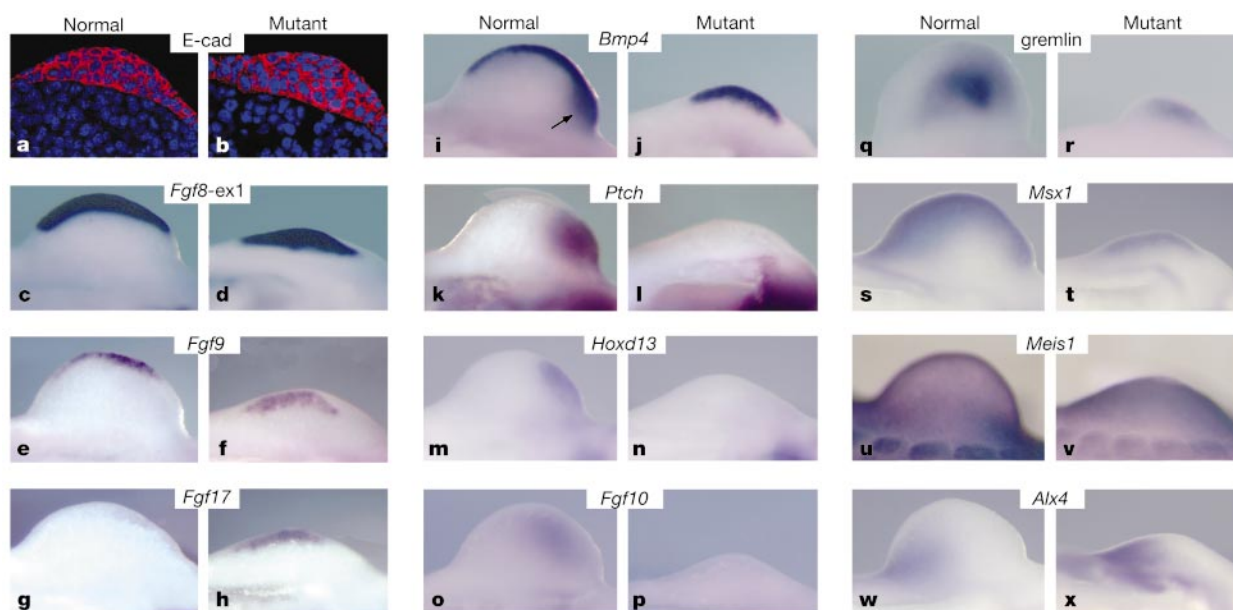


Figure 3 AER morphology and gene expression in hindlimb buds. **a, b**, Transverse sections through the AER of limb buds at 34 som, stained with anti-E-cadherin (red) and Hoechst dye (blue) to illustrate AER morphology. **c–x**, Gene expression assayed by RNA *in situ* hybridization in whole mount. **c–h**, AER-FGF genes: *Fgf8* (35 som), using an *Fgf8* exon-1-specific complementary DNA probe containing sequences not deleted by Cre (320 base pairs (bp) including 84 bp of 5' untranslated region); *Fgf9* (36 som); *Fgf17* (34

som). **i–p**, Genes whose expression was not detected in mutant mesenchyme: *Bmp4* (36 som); *Ptch* (34 som); *Hoxd13* (36 som); *Fgf10* (35 som). The arrow in **i** points to the mesenchymal expression domain of *Bmp4* in the normal limb bud. **q–x**, Genes whose expression was detected in mutant mesenchyme: *gremlin* (40 som); *Msx1* (35 som); *Meis1* (34 som); *Alx4* (35 som).

mesenchyme at 28, 30 and 32 som (Fig. 5c–f, and data not shown). No statistically significant differences were found in the frequency of mitotic cells in the hindlimb bud distal mesenchyme of double knockout compared with normal embryos at any of the three stages quantified (see Supplementary Information). The mitotic index, which ranged from 4.4% to 6.8%, was similar to the values reported for chick forelimb bud mesenchyme at Hamburger–Hamilton stages 21–25 (ref. 31). Thus, loss of *Fgf4* and *Fgf8* function in the AER had no detectable effect on cell proliferation in the underlying mesenchyme. A similar conclusion was drawn from an analysis of 5-bromodeoxyuridine (BrdU) incorporation (not shown). However, we cannot rule out the possibility that cell proliferation is reduced in double knockout hindlimb buds at late stages (>35 som) of development.

Nascent double knockout hindlimbs are abnormally small

We found that considerably before the onset of apoptosis in proximal mesenchyme at 34 som (Fig. 4c, d), double knockout hindlimb buds were smaller than normal (compare Fig. 5c and e). At 28 som, when the hindlimb bud first becomes apparent, they were only ~75% of normal size ($2.8 \pm 0.2 \mu\text{m}^2$ compared with

$3.7 \pm 0.2 \mu\text{m}^2 (\times 10^4)$), as determined by quantifying limb bud area in transverse sections. This size reduction indicates a decrease in cell number as there was no obvious effect on cell density. It must be due to inactivation of *Fgf8* because it occurs when *Fgf8* but not *Fgf4* is normally expressed. Consistent with this conclusion, *Fgf8* single knockout and *Fgf4; Fgf8* double knockout hindlimb buds are equally small at early stages (not shown). Furthermore, we found that double knockout forelimb buds were normal in size at early stages (Fig. 5g, h), and did not begin to show a size decrease relative to normal until about 28 som, shortly after proximal cell death was first detected. Thus, a decrease in limb bud size is observed when *Fgf8* is inactivated before it is expressed (in hindlimb buds) but not when it is transiently expressed at early stages (in forelimb buds). These data provide strong evidence for a previously unknown function of FGF8 in influencing the number of cells present when limb budding first becomes apparent (28 som). Thus, our data suggest that double knockout hindlimb buds are smaller than normal because they start out smaller, and then decrease further in size as a consequence of proximal cell death. In contrast, double knockout forelimb buds are smaller primarily because of proximal cell death.

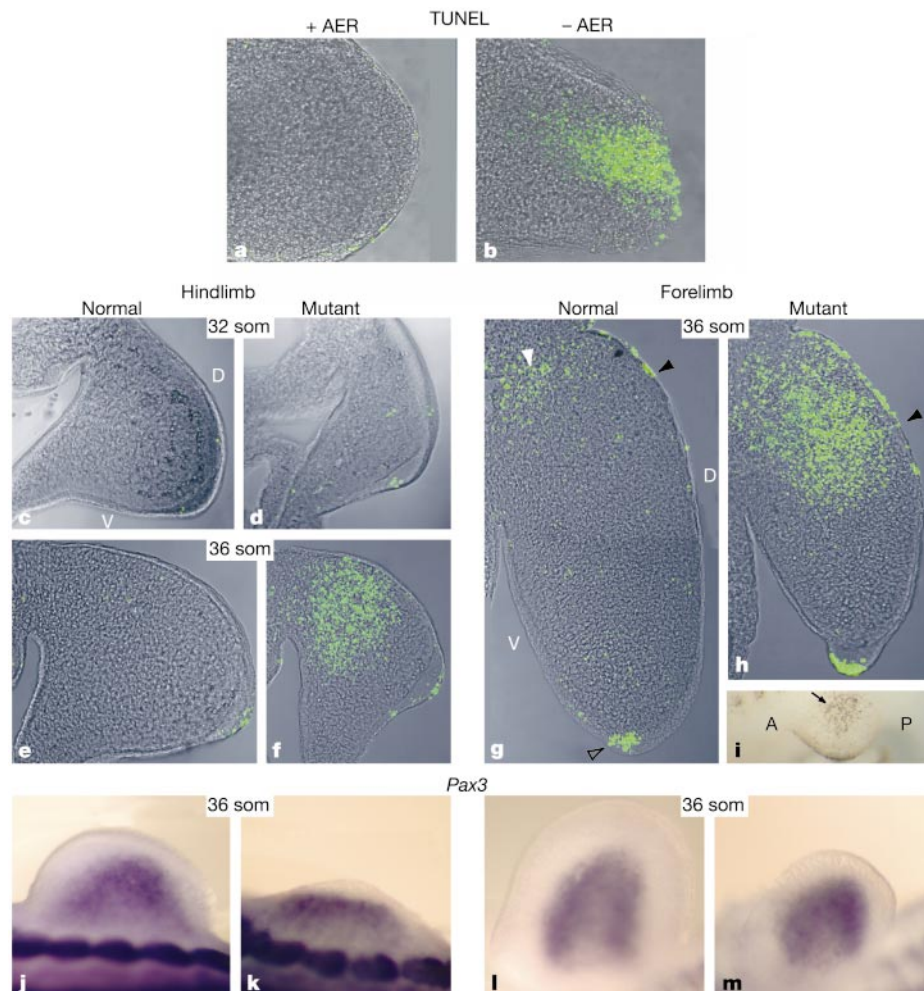


Figure 4 Cell death and distribution of muscle cell precursors. **a–h**, TUNEL assays for cell death in transverse sections of limb buds at the somite stages indicated. In each panel, optical sections containing apoptotic cells (green) are superimposed on a bright field image of the limb bud. **a, b**, Limb buds from a normal ~E10 mouse, cultured *in vitro* for 7 h with the AER intact (**a**) or surgically removed (**b**). **c–h**, Normal and double knockout limb buds. Domains of normal cell death at 36 som are indicated by arrowheads in the proximal mesenchyme (white arrowhead), dorsal ectoderm (black arrowhead), and AER

(open arrowhead). Note the extended domain of cell death in the dorsal ectoderm (black arrowhead) of the mutant limb bud in **h**. **i**, TUNEL assay in whole mount of a 30 som mutant forelimb bud (dorsal view), showing the location of the mesenchymal cell death domain (brown stain, arrow) with respect to the anterior–posterior axis. Cell death is also detected in the AER. **j–m**, Assay for *Pax3* RNA by whole-mount *in situ* hybridization detects muscle precursor cells in limb buds (36 som). A, anterior; D, dorsal; P, posterior; V, ventral.

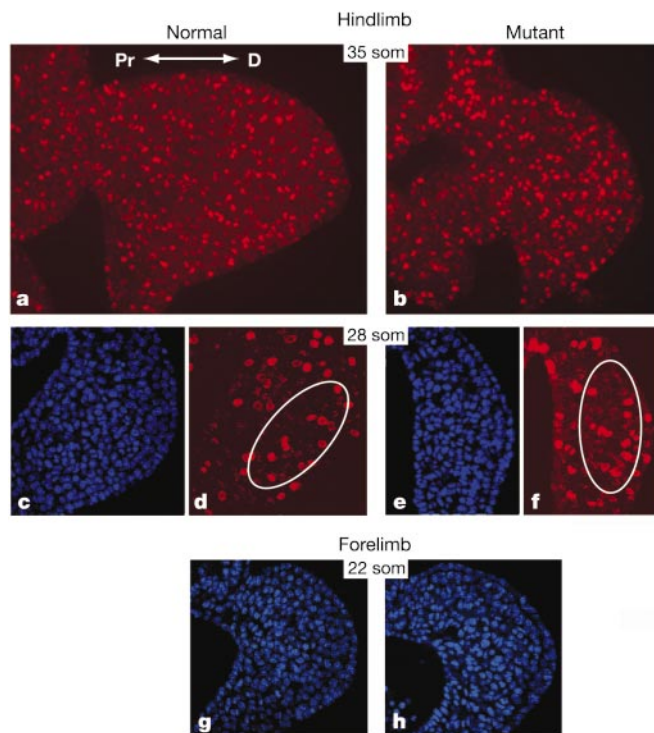


Figure 5 Cell proliferation and limb bud size. **a–h**, Transverse sections of limb buds at the stages indicated, in which phosphorylated histone H3 marks nuclei in cells undergoing mitosis in red and Hoechst labels all nuclei blue. Proximal (Pr) is on the left and distal (D) on the right. Ovals (**d**, **f**) delineate the areas of distal mesenchyme in which mitosis was quantified.

A model for AER-FGF function

Our finding that double knockout embryos lack hindlimbs provides formal genetic evidence that FGF4 and FGF8 together are required to form a limb. In their absence, the other AER-FGFs (FGF9 and FGF17) cannot rescue limb development. Molecular analysis of mutant hindlimb buds, in which functional *Fgf4* and *Fgf8* RNAs are never expressed, demonstrated that these AER-FGFs are not required for AER formation, but are essential for expression of numerous genes in the underlying mesenchyme. For example, without FGF4 and FGF8, we never detected *Shh* expression. However, failure to activate SHH signalling cannot be the sole cause of hindlimb aplasia in our mutants, because *Shh*^{−/−} hindlimbs contain elements representing all three limb segments^{32,33}. Interestingly, when *Shh* expression is induced by transient exposure to FGF8 and FGF4, as in our double knockout forelimb buds, it is maintained, albeit at a much lower level than normal (supplemental data), presumably by FGF9 and FGF17.

Double knockout forelimb buds develop into limbs with decreased skeletal element size or number. These effects were traced to precondensation stages (Fig. 2i–p). Similar phenotypes are observed when limb bud mesenchymal cell number is experimentally reduced^{34,35}, suggesting that FGF4 and FGF8 are required to ensure that enough chondrocyte progenitors are available to form the complete limb skeleton¹⁷ (Fig. 6).

One way that FGF8 seems to perform this function is by influencing cell number in nascent limb buds. When budding is first evident, double knockout hindlimb buds have significantly fewer cells than normal, presumably due to events occurring between the stages when *Fgf8* expression would normally commence and when the effect is observed. This period may be less than 3 h, making it unlikely that small limb bud size is due to reduced proliferation. Cell death is apparently not responsible, because it was not detected before or during budding. Instead, FGF8-dependent

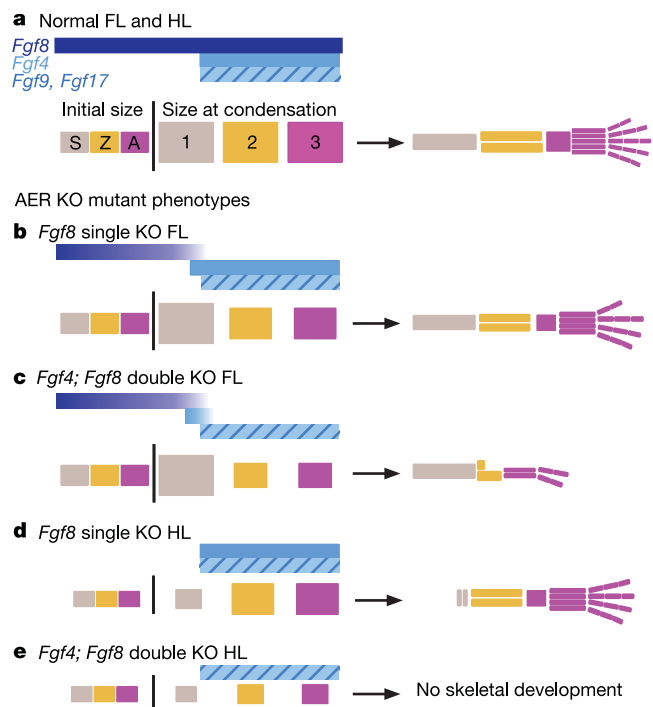


Figure 6 A model for AER-FGF function in the limb. AER-FGF signalling functions before the start of skeletal differentiation to establish the number of chondrocyte progenitors available to form limb skeletal elements¹⁷. We suggest that AER-FGFs affect stylopod (S), zeugopod (Z) and autopod (A) progenitor populations (boxes) in two ways, first by influencing their 'initial size' in the nascent limb bud and subsequently by providing proliferation or survival signals that determine their 'size at condensation' (final size). The model is predicated on two features of limb development. First, progenitor populations of the three limb segments reach their normal sizes in sequence (stylopod (1), zeugopod (2) and then autopod (3))²⁷. Second, expression of *Fgf8* (dark blue) commences as many as 10 somite stages before *Fgf4* (light blue) and *Fgf9* and *Fgf17* (hatched). **a**, In normal limb buds, the initial sizes of all three progenitor populations (arbitrarily shown as equal) are influenced by FGF8 signalling via its effect on nascent limb bud size. The S progenitor population attains final size shortly after expression of other AER-FGFs commences, and therefore is mostly affected by FGF8. Next, Z and then A progenitor populations attain their final sizes during stages when all AER-FGFs are active. For simplicity, the final sizes of all three populations are shown as being equal, although fate-mapping studies indicate that they differ considerably⁴⁸. **b**, **c**, In both *Fgf8* single¹⁷ and *Fgf4*; *Fgf8* double knockout (KO) forelimb (FL) buds, early transient expression of functional *Fgf8* RNA ensures that the initial sizes of the S, Z and A populations and the final size of the S progenitor population are normal; however, the final sizes of Z and A progenitor populations depend on whether FGF4 is present after *Fgf8* is inactivated. In **b**, FGF4, FGF9 and FGF17 signalling substantially rescues Z and A progenitor populations, resulting in slight hypoplasia of Z and A elements, including loss of one digit. In **c**, there is no FGF4 and the phenotype is consequently more severe (Z and A segments are severely hypoplastic and some elements are missing). **d**, In contrast, in *Fgf8* single KO hindlimb (HL) buds¹⁷, the initial sizes of S, Z and A progenitor populations and the final size of the S progenitor population are much smaller than normal owing to absence of FGF8 signalling, resulting in a severely hypoplastic S. Despite their small initial size, Z and A progenitor populations are substantially rescued by FGF4, FGF9 and FGF17, such that Z and A are only mildly hypoplastic and only one digit is absent. **e**, In double KO HL buds, lack of FGF8 and FGF4 signalling affects both initial and final sizes of all progenitor populations. Although there is some rescue of Z and A progenitors by FGF9 and FGF17, and chondrocyte differentiation commences, too few cells are available to form any skeletal elements, consistent with data indicating that reduction below a threshold number results in failure of skeletal development³⁵.

morphogenetic movements or changes in cell–cell adhesion may affect the number of cells that contribute to the limb bud. This hypothesis is consistent with data suggesting that FGF signalling regulates cell movement and cell adhesion in limb buds^{36,37}.

AER-FGF signalling must also be required at later stages, because *Fgf8* single knockout forelimb buds, in which *Fgf8* is inactivated after outgrowth commences¹⁷, form limbs with skeletal hypoplasia. Moreover, when *Fgf4*—which normally commences expression after the limb bud is well established—is also inactivated, forelimb phenotype severity increases. Reduced cell proliferation is unlikely to be responsible for the skeletal phenotypes observed, because cell proliferation appeared similar in mutant and normal limb buds. Instead, the extensive apoptosis in a proximal–dorsal region of double knockout forelimb buds may explain the hypoplasia and aplasia of forelimb zeugopod elements, if their progenitors are among the dying cells. The loss of autopod elements could be an indirect effect if autopod progenitors are recruited into the zeugopod progenitor population to compensate for depletion caused by apoptosis. Thus, death of some progenitors and a fate change in others might account for loss of both proximal and distal elements. This hypothesis is consistent with studies showing that excision of prospective stylopod and zeugopod from chick limb buds can result in absence not only of proximal elements, but digits as well³⁸. Other tissue deficits in the mutant limb buds might have a similar aetiology.

AER-FGFs and limb patterning

An important issue is how our results fit with previously proposed models for limb patterning. They do not appear to be entirely consistent with the progress zone model³⁹, which has been the prevailing model of limb proximal–distal patterning for ~30 years. It postulates that limb progenitors in a ‘progress zone’ in the limb bud distal tip become progressively ‘distalized’. As cells proliferate but progress zone size remains constant, cells must be continually exiting the progress zone. Once they leave, they cease distalizing. Therefore, cells exiting early form proximal skeletal elements whereas those exiting late form distal elements. AER-FGFs are thought to be the permissive factors that allow progress zone cells to continue distalizing, because FGF beads can rescue distal structures after AER removal^{7,8}. It has further been suggested that a function of AER-FGFs in the proposed distalization process is to repress expression of *Meis1* and *Meis2*, which are thought to code for proximalizing factors^{40–42}. Our data are consistent with this hypothesis in that the *Meis1* expression domain occupies a larger proportion of double knockout limb buds than normal (Fig. 3u, v, and Supplementary Information). The presence of one or two distally complete digits in our double knockout forelimbs (Fig. 2c–e) can be explained in terms of the progress zone model by assuming that in the absence of both FGF4 and FGF8, FGF9 and FGF17, which are still produced in mutant AER, might provide the factors necessary for distalization.

Explaining our finding—that proximal elements are invariably missing or severely hypoplastic when distal elements are present—in terms of a model that postulates a progressive process whereby distal is specified only after proximal, is problematic. One could argue that if progress zone cells are killed, distal elements form when proximal ones are absent because the surviving cells remain in the progress zone for longer than normal while replacing those that died. Therefore, they undergo excessive distalization and fail to make proximal elements³⁵. However, as we detected no distal cell death in our mutant limb buds, possibly because *Fgf9* and *Fgf17* are expressed, that explanation does not apply. If it proves true, as we suggested above, that there is a change in cell fate from distal to proximal in the mutant forelimbs, this would argue against the progress zone model (see ref. 27).

At present, we favour a different concept, which fits well with our model of AER-FGF function. It assumes that limb skeletal patterning is achieved as a function of basic cellular processes including cell division, cell survival, and stereotypic behaviours of chondrocyte progenitors such as aggregate formation⁴³. Size and shape of the domains from which cells are recruited to aggregates may determine

the dimensions of emerging condensations. Their final form, and thus the skeletal elements into which they develop, may subsequently be fine-tuned by local cell–cell interactions and signalling^{44,45}. □

Methods

Production and analysis of double knockout mice

Animals with the genotype shown in Fig. 1a were produced by mating *Fgf4*^{flax/flax}; *Fgf8*^{flax/flax} females with *Mx2-cre*; *Fgf4*^{Δ2,3/flax}; *Fgf8*^{flax/+} or *Mx2-cre*/*Mx2-cre*; *Fgf4*^{Δ2,3/+}; *Fgf8*^{Δ2,3/+} males. Offspring were genotyped using DNA isolated from amniotic membrane, head, or tail tissue as a substrate for polymerase chain reaction (PCR), with the following primer pairs: for *Cre*, 5′-TGATGAGGTTTCGCAAGAACC-3′ and 5′-CCATGAGTGAACGAACCTGG-3′; for *Fgf4*, 5′-TCTGGAGAGGAAGTAGGAATGG-3′ and 5′-GAAGAGAAGCAGGCAGATGC-3′; for *Fgf8*, 5′-CTGCAGAACGCCAAGTACG-3′ and 5′-AGCTCCCGCTGGATTCCTC-3′. Approximately 25% of the offspring were found to have the desired mutant genotype. All other offspring, with the exception of those carrying *Mx2-cre*; *Fgf4*^{flax/+}; *Fgf8*^{Δ2,3/flax}, had normal limbs and were used as controls in our analysis. Because some of the control animals carried *Mx2-cre*, we can rule out *cre* expression as a cause of abnormalities detected in double knockout limb buds. We staged embryos by counting somites as previously described¹¹. We performed whole-mount RNA *in situ* hybridization according to a standard protocol.

TUNEL analysis and immunofluorescence assays

Limb buds were fixed in 4% paraformaldehyde and embedded in 4% low melt agarose, then sectioned on a Leica vibrating microtome. Transverse sections (50-μm thick slices) were collected and assayed by TUNEL (Roche) according to the manufacturer's protocol. To stain all nuclei, limb bud slices were incubated in 50 μg ml⁻¹ Hoechst in PBT (PBS with 0.1% Tween) for 10 min. To detect FGF8 protein, a mouse anti-FGF8 monoclonal antibody (R&D Systems), biotinylated anti-mouse immunoglobulin-γ (IgG) antibody (Vector Labs) and fluorescein avidin D (Vector Labs) were used. To detect cells in mitosis, a rabbit anti-phosphorylated histone H3 antibody (Upstate Biotechnology) and Cy3-conjugated goat anti-rabbit IgG antibody (Jackson Lab) were used. To detect E-cadherin, a rat anti-E-cadherin antibody (Sigma) and Cy3-conjugated goat anti-rat IgG antibody (Jackson Lab) were used. Nonspecific antibody binding was blocked by incubating the limb bud slices for 1 h at room temperature in PBT with 5% normal goat serum (NGS, Sigma). All staining reactions were performed in PBT with 1% NGS, and incubated for 1 h at room temperature, except for incubation with primary antibody, which was carried out overnight at 4 °C. Sections were washed in three changes of PBT between each step of the protocol. Stained limb bud slices were mounted in Vectorshield mounting medium and analysed using either a Zeiss Axiophot fluorescence microscope or a Leica TCS NT confocal microscope.

AER removal experiments

CD1 mouse embryos (Charles River Laboratories) at about E10 were dissected in cold PBS, and limb-forming regions were isolated by two transverse cuts through the trunk, one above and one below the level of the forelimb buds. After evisceration and removal of the entire AER from one limb bud with fine forceps and tungsten needles, the embryo fragments were cultured in a rotating apparatus (BTC Engineering) for a total of 7 h in 75% rat serum and 25% Bigger's Medium (BGJb Fitton–Jackson modification) with 20% O₂, 5% CO₂ and 75% N₂.

Received 1 November 2001; accepted 5 June 2002; doi:10.1038/nature00902.

- Hinchliffe, J. R. & Johnson, D. R. *The Development of the Vertebrate Limb: An Approach Through Experiment, Genetics, and Evolution* (Clarendon, Oxford, 1980).
- Johnson, R. L. & Tabin, C. J. Molecular models for vertebrate limb development. *Cell* **90**, 979–990 (1997).
- Saunders, J. W. Jr The proximo-distal sequence of the origin of the parts of the chick wing and the role of the ectoderm. *J. Exp. Zool.* **108**, 363–403 (1948).
- Summerbell, D. A quantitative analysis of the effect of excision of the AER from the chick limb bud. *J. Embryol. Exp. Morphol.* **32**, 651–660 (1974).
- Rowe, D. A. & Fallon, J. F. The proximodistal determination of skeletal parts in the developing chick leg. *J. Embryol. Exp. Morphol.* **68**, 1–7 (1982).
- Rubin, L. & Saunders, J. W. J. Ectodermal-mesodermal interactions in the growth of limb buds in the chick embryo: constancy and temporal limits of the ectodermal induction. *Dev. Biol.* **28**, 94–112 (1972).
- Niswander, L., Tickle, C., Vogel, A., Booth, I. & Martin, G. R. FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell* **75**, 579–587 (1993).
- Fallon, J. et al. FGF-2: Apical ectodermal ridge growth signal for chick limb development. *Science* **264**, 104–107 (1994).
- Ornitz, D. M. & Itoh, N. Fibroblast growth factors. *Genome Biol.* **2**, 3005.1–3005.12 (2001).
- Martin, G. R. The roles of FGFs in the early development of vertebrate limbs. *Genes Dev.* **12**, 1571–1586 (1998).
- Sun, X. et al. Conditional inactivation of *Fgf4* reveals complexity of signalling during limb bud development. *Nature Genet.* **25**, 83–86 (2000).
- Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. & Goldfarb, M. Requirement of FGF-4 for postimplantation mouse development. *Science* **267**, 246–249 (1995).
- Sun, X., Meyers, E. N., Lewandoski, M. & Martin, G. R. Targeted disruption of *Fgf8* causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev.* **13**, 1834–1846 (1999).
- Colvin, J. S., Green, R. P., Schmahl, J., Capel, B. & Ornitz, D. M. Male-to-female sex reversal in mice lacking fibroblast growth factor 9. *Cell* **104**, 875–889 (2001).
- Xu, J., Liu, Z. & Ornitz, D. M. Temporal and spatial gradients of *Fgf8* and *Fgf17* regulate proliferation

- and differentiation of midline cerebellar structures. *Development* **127**, 1833–1843 (2000).
16. Moon, A. M., Boulet, A. M. & Capecchi, M. R. Normal limb development in conditional mutants of Fgf4. *Development* **127**, 989–996 (2000).
17. Lewandoski, M., Sun, X. & Martin, G. R. Fgf8 signalling from the AER is essential for normal limb development. *Nature Genet.* **26**, 460–463 (2000).
18. Moon, A. M. & Capecchi, M. R. Fgf8 is required for outgrowth and patterning of the limbs. *Nature Genet.* **26**, 455–459 (2000).
19. Meyers, E. N., Lewandoski, M. & Martin, G. R. An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nature Genet.* **18**, 136–141 (1998).
20. Wright, E. *et al.* The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nature Genet.* **9**, 15–20 (1995).
21. Min, H. *et al.* Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to *Drosophila* branchless. *Genes Dev.* **12**, 3156–3161 (1998).
22. Sekine, K. *et al.* Fgf10 is essential for limb and lung formation. *Nature Genet.* **21**, 138–141 (1999).
23. Revest, J. M. *et al.* Fibroblast growth factor receptor 2-IIIb acts upstream of Shh and Fgf4 and is required for limb bud maintenance but not for the induction of Fgf8, Fgf10, Msx1, or Bmp4. *Dev. Biol.* **231**, 47–62 (2001).
24. Prahlad, K. V., Skala, G., Jones, D. G. & Briles, W. E. Limbless: a new genetic mutant in the chick. *J. Exp. Zool.* **209**, 427–434 (1979).
25. Takahashi, M. *et al.* The role of Alx-4 in the establishment of anteroposterior polarity during vertebrate limb development. *Development* **125**, 4417–4425 (1998).
26. Rowe, D. A., Cairns, J. M. & Fallon, J. F. Spatial and temporal patterns of cell death in limb bud mesoderm after apical ectodermal ridge removal. *Dev. Biol.* **93**, 83–91 (1982).
27. Dudley, A. T., Ros, M. A. & Tabin, C. J. A re-examination of proximodistal patterning during vertebrate limb development. *Nature* **418**, 539–544 (2002).
28. Bober, E., Franz, T., Arnold, H. H., Gruss, P. & Tremblay, P. Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. *Development* **120**, 603–612 (1994).
29. Baldwin, H. S. *et al.* Platelet endothelial cell adhesion molecule-1 (PECAM-1/CD31): alternatively spliced, functionally distinct isoforms expressed during mammalian cardiovascular development. *Development* **120**, 2539–2553 (1994).
30. Gurley, L. R., D'Anna, J. A., Barham, S. S., Deaven, L. L. & Tobey, R. A. Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells. *Eur. J. Biochem.* **84**, 1–15 (1978).
31. Hornbruch, A. & Wolpert, L. Cell division in the early growth and morphogenesis of the chick limb. *Nature* **226**, 764–766 (1970).
32. Chiang, C. *et al.* Manifestation of the limb prepattern: limb development in the absence of Sonic hedgehog function. *Dev. Biol.* **236**, 421–435 (2001).
33. Kraus, P., Fraidenreich, D. & Loomis, C. A. Some distal limb structures develop in mice lacking Sonic hedgehog signalling. *Mech. Dev.* **100**, 45–58 (2001).
34. Alberch, P. & Gale, E. A. Size dependence during the development of the amphibian foot. Colchicine-induced digital loss and reduction. *J. Embryol. Exp. Morphol.* **76**, 177–197 (1983).
35. Wolpert, L., Tickle, C. & Sampford, M. The effect of cell killing by X-irradiation on pattern formation in the chick limb. *J. Embryol. Exp. Morphol.* **50**, 175–193 (1979).
36. Li, S. & Muneoka, K. Cell migration and chick limb development: chemotactic action of FGF-4 and the AER. *Dev. Biol.* **211**, 335–347 (1999).
37. Saxton, T. M. *et al.* The SH2 tyrosine phosphatase Shp2 is required for mammalian limb development. *Nature Genet.* **24**, 420–423 (2000).
38. Stark, R. J. & Searls, R. L. The establishment of the cartilage pattern in the embryonic chick wing, and evidence for a role of the dorsal and ventral ectoderm in normal wing development. *Dev. Biol.* **38**, 51–63 (1974).
39. Summerbell, D., Lewis, J. H. & Wolpert, L. Positional information in chick limb morphogenesis. *Nature* **244**, 492–496 (1973).
40. Capdevila, J., Tsukui, T., Rodriguez Esteban, C., Zappavigna, V. & Izpisua Belmonte, J. C. Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. *Mol. Cell* **4**, 839–849 (1999).
41. Mercader, N. *et al.* Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature* **402**, 425–429 (1999).
42. Mercader, N. *et al.* Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127**, 3961–3970 (2000).
43. Maini, P. K. & Solursh, M. Cellular mechanisms of pattern formation in the developing limb. *Int. Rev. Cytol.* **129**, 91–133 (1991).
44. Dahn, R. D. & Fallon, J. F. Interdigital regulation of digit identity and homeotic transformation by modulated BMP signalling. *Science* **289**, 438–441 (2000).
45. Kingsley, D. M. Genetic control of bone and joint formation. *Novartis Found. Symp.* **232**, 213–222 (2001).
46. Muneoka, K., Wanek, N. & Bryant, S. V. Mammalian limb bud development: *in situ* fate maps of early hindlimb buds. *J. Exp. Zool.* **249**, 50–54 (1989).
47. Carter, T. The genetics of luxate mice. IV. Embryology. *J. Genet.* **52**, 1–35 (1954).
48. Vargesson, N. *et al.* Cell fate in the chick limb bud and relationship to gene expression. *Development* **124**, 1909–1918 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank M. Lewandoski, whose studies on the phenotype of *Fgf8* single mutant embryos first suggested the model of AER-FGF function described here, for discussion at the outset of this work. We are grateful to J. Fallon, D. Kingsley and C. Tickle for discussion, and C. Tabin and A. Dudley for sharing unpublished information and helpful comments. We also thank D. Duboule, P. Gruss, R. Harland, B. Hogan, N. Itoh, J.-C. Izpisua-Belmonte, V. Lefebvre, A. McMahon, D. Ornitz, P. Sharpe, M. Scott and R. Wisdom for providing plasmids from which probes were prepared; D. Lakeland for help with statistical analysis; C. Larabell for technical advice on confocal microscopy; C. Petersen, Z. Serrano, J. Watanabe and R. Rantala for technical assistance; and H. Ingraham, J. Saunders, G. Schubiger and our colleagues in the Martin laboratory for comments on the manuscript. X.S. was the recipient of a postdoctoral fellowship from the American Cancer Society and was supported by an NIH Training grant. F.M. is the recipient of an individual National Research Service Award from the NIH. This work was supported by an HHMI Research Resources Program grant to the UCSF School of Medicine, and an NIH grant to G.R.M.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.R.M. (e-mail: gmartin@itsa.ucsf.edu).

Ultrafast precessional magnetization reversal by picosecond magnetic field pulse shaping

Th. Gerrits*, H. A. M. van den Berg*, J. Hohlfeld*, L. Bär† & Th. Rasing*

* Research Institute for Materials, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

† Siemens AG, CTMF 1, Paul-Gossen-Strasse 100, 91052 Erlangen, Germany

Since the invention of the first magnetic memory disk in 1954, much effort has been put into enhancing the speed, bit density and reliability of magnetic memory devices. In the case of magnetic random access memory (MRAM) devices, fast coherent magnetization rotation by precession of the entire memory cell is desired^{1–6}, because reversal by domain-wall motion is much too slow. In principle, the fundamental limit of the switching speed via precession is given by half of the precession period. However, under-critically damped systems exhibit severe ringing^{7,8} and simulations show that, as a consequence, undesired back-switching of magnetic elements of an MRAM can easily be initiated by subsequent write pulses, threatening data integrity. We present a method to reverse the magnetization in under-critically damped systems by coherent rotation of the magnetization while avoiding any ringing. This is achieved by applying specifically shaped magnetic field pulses that match the intrinsic properties of the magnetic elements. We demonstrate, by probing all three magnetization components^{9,10}, that reliable precessional reversal in lithographically structured micrometre-sized elliptical permalloy elements is possible at switching times of about 200 ps, which is ten times faster than the natural damping time constant.

Changing the binary state of a storage element implies a reversal of its magnetization, which can be achieved by an external field parallel to the magnetization (**M**). In that case, the switching of **M** is accomplished by domain-wall-motion, which is a relatively slow process compared to a reversal by coherent rotation. Landau and Lifshitz gave a phenomenological description of this dynamics¹¹:

$$\frac{d\mathbf{M}}{dt} = -\gamma(\mathbf{M} \times \mathbf{H}_{\text{eff}}) - \lambda[\mathbf{M} \times (\mathbf{M} \times \mathbf{H}_{\text{eff}})] \quad (1)$$

where γ is the gyroscopic constant, representing the precessional frequency. λ is the phenomenological damping factor that drives the system towards an energy minimum. In general, λ is small compared to γ/M_s (where M_s is the saturation magnetization), which implies that the relaxation of the system takes much longer than a few full precessional cycles of **M** (this underdamped behaviour is the so-called ringing). $\mathbf{H}_{\text{eff}}$ is the internal effective field of the system that in our experiments is the sum of the external field $\mathbf{H}_{\text{bias}}$ + $\mathbf{H}_{\text{pulse}}$, the anisotropy $\mathbf{H}_{\text{ani}}$ and the demagnetizing field $\mathbf{H}_{\text{dem}}$. The field produced by the anisotropy points in the direction of the long axis of the element. The demagnetizing field depends on the direction of **M**. With **M** saturated along one of the in-plane axes of our element, the in-plane components of $\mathbf{H}_{\text{dem}}$ are of the order of a few gauss, being small compared to the other involved in-plane fields. Contrary to this, the vertical component of $\mathbf{H}_{\text{dem}}$ ($= -4\pi M_z$) is large (of the order of 10^2 G). H_z^{dem} forces the magnetization to rotate in the film plane and determines the actual speed of the precessional motion. Recent experiments^{7,12,13} show that it is possible to stop a magnetization precession without oscillating around the equilibrium direction, at which **M** is in the thin-film plane, and M_z and consequently H_z^{dem} are equal to zero. With a field pulse of rectangular course in time with given maximum amplitude, the fastest possible switching can be accomplished by properly adjusting the pulse time, so that the motion will be stopped at half the

precession period. However, the shape of the real individually generated field pulses, as in the present experiment, is fixed by the time constants of the leading and trailing edge. The required second degree of freedom for the stopping of **M** can be obtained by superimposing a second field, the quenching field, with the appropriate delay and strength, so that the initially generated field is compensated at the trailing edge of the quenching field. This compensation can be almost perfect, owing to the exponential character, dominated by one time constant only, of the decay at the trailing edge. The quenching field is used to align $\mathbf{H}_{\text{eff}}$ parallel to **M** at the moment that **M** is along a state of static equilibrium, resulting in a torque that is equal to zero, so that the motion terminates.

The desired shape of the magnetic field in time is obtained by using two independently triggered GaAs photoswitches, excited by femtosecond laser pulses^{14–19}. A schematic drawing of this approach is shown in Fig. 1. A laser pulse generates a charge package inside a photoconductive switch through which a current package is transferred from the electrode to the strip line. The shape of this current package is determined by the photoswitch characteristics⁵. The associated magnetic field propagates down a signal line that is embedded in a coplanar waveguide structure. For shaping this field in time, a second photoswitch is implemented that launches an additional current package. The second current package, which is superimposed on the first, consequently modifies the magnetic field. The sign and amplitude of the voltages at the switches and the time delay between the two exciting laser pulses are the available parameters. Experiments showed that the precessional motion could be stopped after one period by superimposing the second generated current package onto the first; cancellation of the total field perpendicular to **M** was then achieved¹³. The experiments reported here were performed with a commercial Ti:sapphire laser system, operating at 800 nm, and at 76 MHz repetition rate with a pulse width of 150 fs. The magnetic response is stroboscopically reconstructed by a focused probe-laser pulse that probes the magnetization at a distinct time delay after arrival of the (shaped) magnetic field. This time-delay can be varied by using a delay-line that can alter the optical path between the pump and the probe

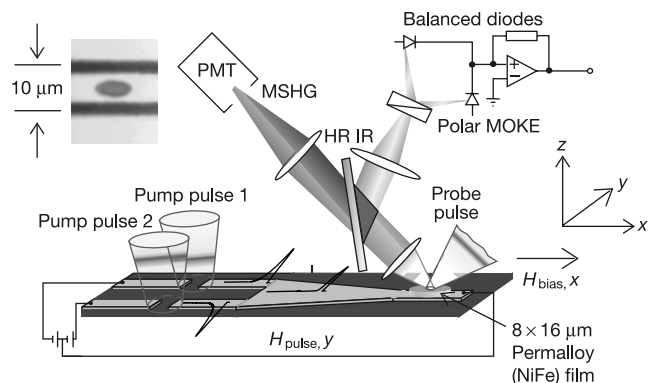


Figure 1 The schematic experimental set-up. A sequence of two pump-laser-pulses, which are separated by a distinct time-delay, excites two GaAs-photoconductive switches. The generated current pulses are superimposed to produce one short and square-like magnetic field pulse. The field pulse is launched down a coplanar waveguide structure and excites the thin-film permalloy ($\text{Ni}_{81}\text{Fe}_{19}$) element at the end of the tapering. The inset is a microscopic photograph of the 8-nm-thin permalloy magnetic element that has an elliptical shape with dimensions of $8 \times 16 \mu\text{m}$. The vector- and time-resolved element response is measured by magnetization-induced second harmonic generation (MSHG) and the polar magneto-optical Kerr effect (MOKE). A high-reflectance infrared mirror (HRIR) is used to split the fundamental and second harmonic part of the beam. A photomultiplier tube (PMT) is used to detect the second harmonic photons.

pulses. The stroboscopic, time-resolved experiment requires a recovery of the initial state before each pump-probe sequence, for which purpose an external bias-field is applied. Moreover, measurement of all magnetization components is necessary for resolving the magnetization dynamics completely, so that coherence can be verified. Therefore, the in-plane components of $\mathbf{M}$ are measured by magnetization-induced second harmonic generation (MSHG)^{20,21}. The linear polar magneto-optical Kerr effect (MOKE) is used to probe the polar component of $\mathbf{M}$.

The ferromagnetic 8-nm-thin elliptic permalloy film element, with a long and short axis of 16 μm and 8 μm , respectively (inset of Fig. 1), was situated at the top of the signal line (Fig. 1). Static magnetization-loop measurements confirmed that the easy anisotropy-axis of the element lay along the long axis of the element, in accordance with the shape anisotropy and the induced sputter anisotropy. An elliptic shape of the element was chosen in order to achieve a uniform demagnetizing field inside the element during

magnetization reversal. Vibrating sample magnetometer characterization of the thin film gave an absolute value for $\mathbf{M}$ of 800 e.m.u. cm^{-3} . The demagnetization factors of the thin film element were calculated to be: $N_x = 0.0002$, $N_y = 0.0006$ and $N_z = 0.9992$, resulting in demagnetizing fields along the x , y and z axes of 2 oersteds (Oe), 6 Oe and 10 kOe respectively.

To obtain fast switching, a strong excitation-field, an additional quenching field at the moment that M_{polar} is zero and coherence of

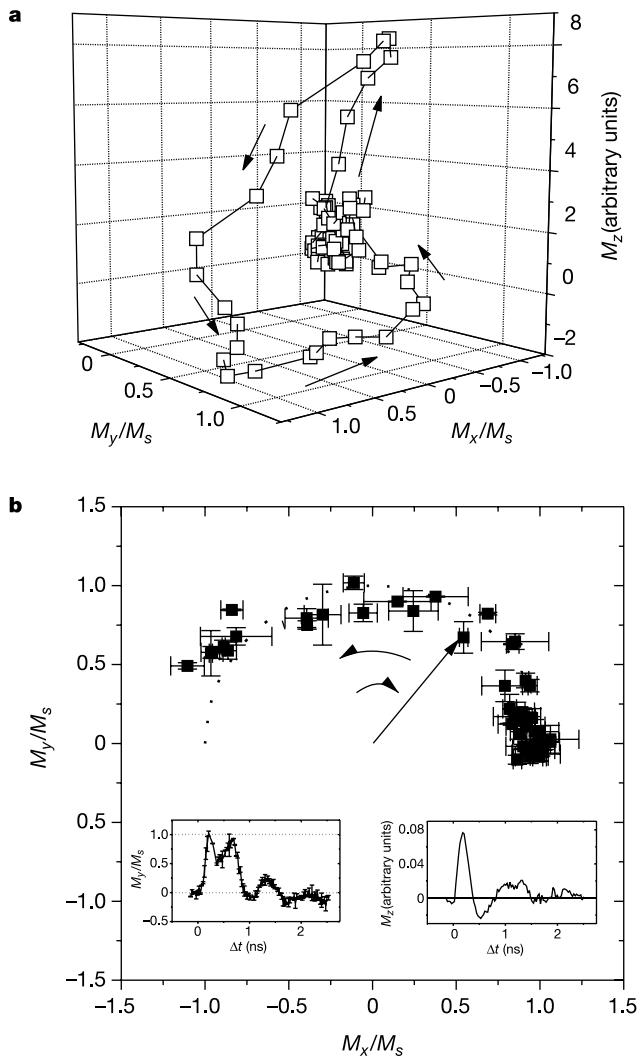


Figure 2 Large field excitation without stopping. **a**, The three-dimensional trajectory of $\mathbf{M}$ while excited with a 70-Oe in-plane field pulse. We note the small polar component of $\mathbf{M}$ is given in arbitrary units and enlarged relative to the in-plane components. **b**, The in-plane projection of $\mathbf{M}$. The two in-plane components were probed by MSHG. A clear coherent 160° reversal can be observed. The ringing effect, indicated by the two insets in **b** representing the y and z components of $\mathbf{M}$ respectively, is present. After maximum excursion, the non-zero torque acting on $\mathbf{M}$ due to the present fields drives $\mathbf{M}$ back to its initial state while showing a pronounced effect of ringing.

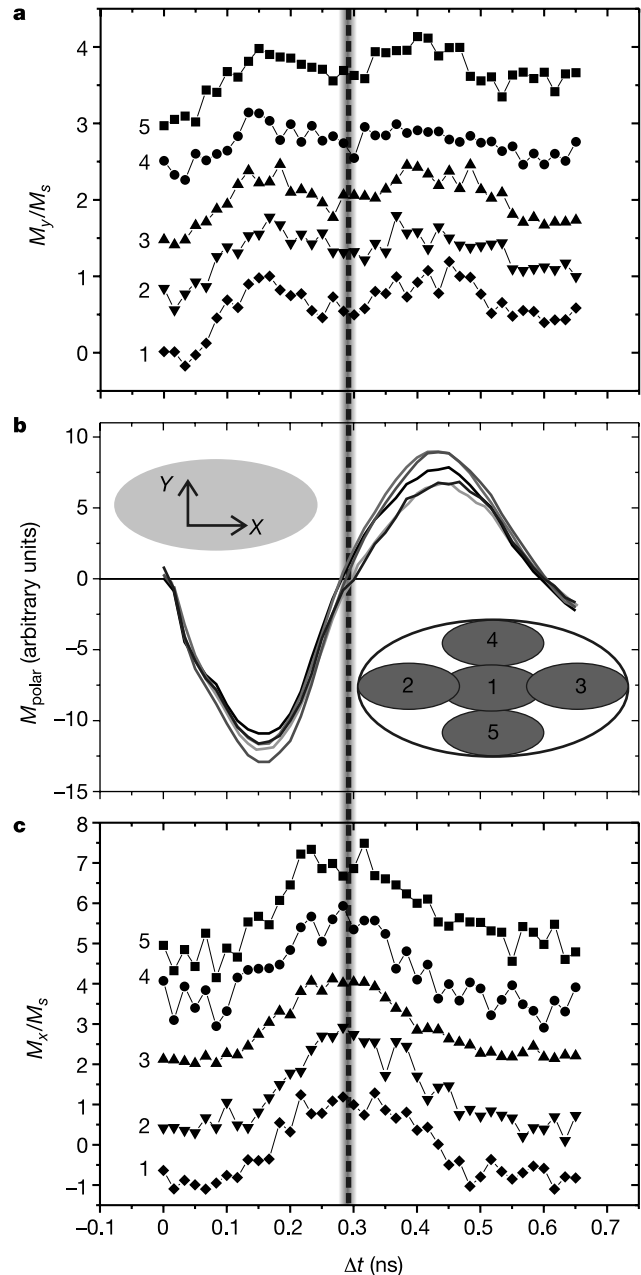


Figure 3 The in-plane and polar magnetic response of the element measured on five different locations to verify coherent rotation of the whole element. The probing area of the laser spot has an elliptical shape due to the beam-waist and the 45° incident beam which is represented by the five ellipses in the bottom right inset of **b**. A high degree of coherence can be seen by the in-plane components (**a**, **c**) that all show the large-angle rotation and by the polar component (**b**) showing the moment when $\mathbf{M}$ passes the film plane for the individual locations. From this we can extract the degree of incoherence (grey vertical bar at 0.29 ns in this figure) which is as low as $\pm 5\%$ relative to the time needed for the maximum in-plane excursion (vertical dashed line). The grey ellipse in the top left inset of **b** shows the axes with respect to the element.

the system are indispensable. We measured the response due to a large magnetic field pulse with 70 Oe pulse-peak, 10 ps rise-time and 400 ps decay-time. Figure 2a shows the trajectory of $\mathbf{M}$. Figure 2b shows the projection of the normalized $\mathbf{M}$ in the plane of the film. This strong single-field-pulse excitation initiates a rotation in the opposite direction, where the $\mathbf{M}$ excursion is as large as 160° , with clear signs of ringing (as seen in the insets of Fig. 2b) due to the fields that are still present at maximum excursion. The ringing effect causes $\mathbf{M}$ to flip back and end up in the initial state. By carrying out vector-resolved measurements at five different locations over the whole element, we could observe a high degree of coherence. The measurements are presented in Fig. 3. Figure 3a and c show the two in-plane components of $\mathbf{M}$ ($x \parallel$ long axis; $y \parallel$ pulse field) which show the large-angle excitation with a maximum excursion of $\mathbf{M}$ after a mean time of 290 ps, indicated by the dashed line. Figure 3b shows the polar component, illustrating that at maximum in-plane excursion, passing of the thin-film plane takes place within an uncertainty of 25 ps, represented by the grey area. From these observations, we can deduce a degree of incoherence which is as low as $\pm 5\%$. This result is also in accordance with recently obtained theoretical results²¹, showing calculations for an elliptical-shaped thin-film element. A high degree of coherence is necessary for a successful suppression of the ringing effect after reversal. The solid dots in Fig. 4 show the response of the magnetic system while launching the quenching field at maximum in-plane excursion from the initial state when M_z is zero. A clear magnetization reversal, followed by stopping of the system's motion and thus suppression of the ringing is observed. This reversal is reached in about 200 ps, which is much faster than the natural damping time constant.

This technique may have a large impact on MRAM devices. A precessional magnetization reversal is fast, and here the writing process already reaches 5 GHz. From equation (1), we conclude that the correct choice of material, field pulse and film parameters may greatly decrease the writing time for a single bit. Furthermore the pulse-shaping technique allows a much larger tolerance regime for the material parameters, because changes can be compensated for electronically.

We have thus demonstrated a method for fast switching of magnetic elements by suppressing the unwanted effect of ringing originating from the precessional motion, by applying specifically shaped picosecond magnetic field pulses. We have demonstrated a

high degree of coherence of $\mathbf{M}$ within the experimental limitations. Application of the pulse-shaping technique on a picosecond time-scale leads to less stringent tolerance windows for magnetic elements. In addition, parameter changes can be compensated by the correct choice of the field pulse. This technique is essential for the reliability of MRAM devices where the only crucial parameter is the shape of the magnetic field pulse on the bit line. Our approach has the additional advantage that the pulses are of single polarity. Thus ultrafast precessional magnetization reversal in under-critically damped systems with switching rates of at least several GHz is possible.

While this paper was being peer reviewed, we learned about other groups that have observed precessional switching through various techniques based on pulse generators evoking the magnetic field pulses^{23–25}. Spatially resolved measurements using easy-axis pulses were used by Hiebert *et al.*²³; Schumacher *et al.*²⁵ and Kaka and Russek²⁴ used spin-valve structures and hard-axis pulses to follow the effect of precessional switching. □

Received 30 January; accepted 31 May 2002; doi:10.1038/nature00905.

- Hiebert, W. K., Stankiewicz, A. & Freeman, M. R. Direct observation of magnetic relaxation in small Permalloy disk by time-resolved scanning Kerr microscopy. *Phys. Rev. Lett.* **79**, 1134–1137 (1997).
- Back, C. H. *et al.* Magnetization reversal in ultrashort magnetic field pulses. *Phys. Rev. Lett.* **81**, 3251–3254 (1998).
- Stamps, R. L. & Hillebrands, B. Biased switching of small magnetic particles. *Appl. Phys. Lett.* **75**, 1143–1145 (1999).
- Choi, B. C., Belov, M., Hiebert, W. K., Ballentine, G. K. & Freeman, M. R. Ultrafast magnetization reversal dynamics investigated by time domain imaging. *Phys. Rev. Lett.* **86**, 728–731 (2001).
- Gerrits, Th. *et al.* Precession dynamics in NiFe thin films, induced by short magnetic in-plane field pulses generated by a photoconductive switch. *J. Magn. Soc. Jpn* **25**, 192–197 (2001).
- Acremann, Y. *et al.* Ultrafast generation of magnetic fields in a Schottky diode. *Nature* **414**, 51–54 (2001).
- Bauer, M., Lopusnik, R., Fassbender, J. & Hillebrands, B. Suppression of magnetic-field pulse-induced magnetization precession by pulse tailoring. *Appl. Phys. Lett.* **76**, 2758–2760 (2000).
- Acremann, Y., Back, C. H., Buess, M., Pescia, D. & Pokrovsky, V. Bifurcation in precessional switching. *Appl. Phys. Lett.* **79**, 2228–2230 (2001).
- Acremann, Y. *et al.* Imaging precessional motion of the magnetization vector. *Science* **290**, 492–494 (2000).
- Silva, T. J., Pufall, M. R. & Kabos, P. Nonlinear magneto-optic measurement of flux propagation dynamics in thin Permalloy films. *J. Appl. Phys.* **91**, 1066–1073 (2002).
- Landau, L. & Lifshitz, E. On the theory of dispersion of magnetic permeability in ferromagnetic bodies. *Phys. Z. Union* **8**, 153–169 (1935).
- Crawford, T. M., Kabos, P. & Silva, T. J. Coherent control of precessional dynamics in thin film permalloy. *Appl. Phys. Lett.* **76**, 2113–2115 (2000).
- Gerrits, Th. *et al.* Picosecond control of coherent magnetization dynamics in Permalloy thin films by picosecond magnetic field pulse shaping. *J. Magn. Magn. Mater.* **240**, 283–286 (2002).
- Auston, D. H. Impulse response of photoconductors in transmission lines. *IEEE J. Quantum Electr.* **19**, 639–648 (1983).
- Frankel, M. Y., Gupta, S., Valdmanis, J. A. & Mourou, G. A. Terahertz attenuation and dispersion characteristics of coplanar transmission lines. *IEEE Trans. Microwave Theory Technol.* **39**, 910–915 (1991).
- Keil, U. D. *et al.* High-speed coplanar transmission lines. *IEEE J. Quantum Electr.* **28**, 2333–2342 (1992).
- Keil, U. D., Gerritsen, H. J., Haverkort, J. E. M. & Wolter, H. J. Generation of ultrashort electrical pulses with variable pulse width. *Appl. Phys. Lett.* **66**, 1629–1631 (1994).
- Keil, U. D. & Dykaar, D. R. Ultrafast pulse generation in photoconductive switches. *IEEE J. Quantum Electr.* **32**, 1664–1671 (1996).
- Holzman, J. E., Vermeulen, F. E. & Elezzabi, A. Y. Recombination-independent photogeneration of ultra-short electrical pulses. *Appl. Phys. Lett.* **76**, 134–136 (2000).
- Rasing, Th. Nonlinear magneto-optical probing of magnetic interfaces. *Appl. Phys. B* **68**, 477–484 (1999).
- Crawford, T. M., Silva, T. J., Teplin, C. P. & Rogers, C. T. Subnanosecond magnetization dynamics measured by the second-harmonic magneto-optic Kerr effect. *Appl. Phys. Lett.* **74**, 3386–3388 (1999).
- Zhu, J. & White, M. *Magnetization Reversal and Energy Damping?* (Intermag Europe 2002, Amsterdam, The Netherlands, 2002).
- Hiebert, W. K., Ballentine, G. E. & Freeman, M. R. Comparison of experimental and numerical micromagnetic dynamics in precessional switching and modal oscillations. *Phys. Rev. B* **65**, R140404-1–R140404-4 (2002).
- Kaka, S. & Russek, S. E. Precessional switching of submicrometer spin-valves. *Appl. Phys. Lett.* **86**, 2958–2960 (2002).
- Schumacher, H. W. *et al.* *Precessional Magnetization Reversal in Microscopic Spin Valve Cells* (Intermag Europe 2002, Amsterdam, The Netherlands, 2002).

Acknowledgements

The authors wish to thank all members of CTMF1 at Siemens for their collaboration, and for guidance with preparing the devices. This work was part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie (FOM) and financially

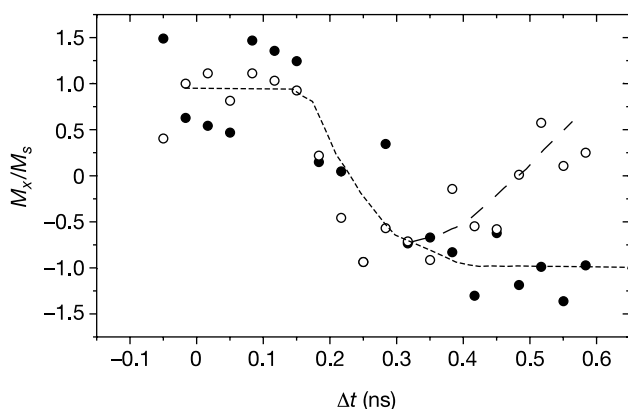


Figure 4 Switching by large-field excitation and suppression of ringing. Without a stop pulse, the system switches back to its initial state (open circles). After sending the stop pulse, the suppression of the ringing of the magnetization can clearly be observed (solid circles). The lines are guides to the eye. The low signal-to-noise ratio in the M_x component results from the very weak longitudinal MSHG signal with an incoming polarization parallel and second harmonic polarization perpendicular to the plane of incidence (p_{in} – s_{out}).

supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), Philips Research and partly supported by the Brite Euram project Tunnel Sense.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Th.R. (e-mail: theoras@sci.kun.nl).

Single-pulse coherently controlled nonlinear Raman spectroscopy and microscopy

Nirit Dudovich, Dan Oron & Yaron Silberberg

Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel

Molecular vibrations have oscillation periods that reflect the molecular structure, and are hence being used as a spectroscopic fingerprint for detection and identification. At present, all nonlinear spectroscopy schemes use two or more laser beams to measure such vibrations¹. The availability of ultrashort (femtosecond) optical pulses with durations shorter than typical molecular vibration periods has enabled the coherent excitation of molecular vibrations using a single pulse². Here we perform single-pulse vibrational spectroscopy on several molecules in the liquid phase, where both the excitation and the readout processes are performed by the same pulse. The main difficulty with single-pulse spectroscopy is that all vibrational levels with energies within the pulse bandwidth are excited. We achieve high spectral resolution, nearly two orders of magnitude better than the pulse bandwidth, by using quantum coherent control techniques. By appropriately modulating the spectral phase of the pulse we are able to exploit the quantum interference between multiple paths to selectively populate a given vibrational level, and to probe this population using the same pulse. This scheme, using a single broadband laser source, is particularly attractive for nonlinear microscopy applications, as we demonstrate by constructing a coherent anti-Stokes Raman (CARS) microscope operating with a single laser beam.

Nonlinear coherent vibrational spectroscopy has evolved during the last two decades as a result of the availability of short pulse sources, the high peak power of which enables detection of the inherently weak nonlinear signals^{1,3}. This includes a family of nonlinear processes, typically involving a third-order nonlinearity (that is, three exciting photons) and a Raman active vibrational level. The most commonly used scheme is CARS, the energy level diagram of which is shown in Fig. 1a. In this process, two photons with frequencies ω_1 and ω_2 excite a vibrational level at an energy $\hbar(\omega_1 - \omega_2)$. A third photon at ω_3 interacts with the excited level to emit a signal photon at a frequency $(\omega_1 + \omega_3 - \omega_2)$. Typically, two narrow-band (narrower than the typical linewidth of Raman levels, corresponding to picosecond pulses) beams are used, exciting only Raman levels in resonance with their energy difference (in this case $\omega_1 = \omega_3$). Time-resolved CARS is a variant in which several Raman levels are simultaneously populated by one or two large bandwidth excitation pulses⁴ and probed by a delayed probe pulse.

This work demonstrates that the entire CARS process can be incorporated within a single ultrashort pulse. All three photons required for the process are supplied by the same short optical pulse,

while quantum coherent control techniques are implemented to restore the high spectral resolution. Furthermore, this technique can overcome the strong nonresonant background which typically accompanies CARS with femtosecond pulses⁵.

In coherent control, constructive or destructive interference is induced between the different paths leading to a final quantum state, to manipulate the outcome of a light-matter interaction^{6–8}. Control of the simplest nonlinear system, a nonresonant two-photon transition between two levels, was demonstrated by Weiner *et al.*⁹ for a Raman transition, and by Meshulach and Silberberg¹⁰ for two-photon absorption. Control over more complicated systems often involves the use of adaptive techniques¹¹, where the spectral phase is determined by a feedback optimization procedure. This approach has been applied to a variety of systems, such as selective chemical bond breaking¹², high-harmonic generation¹³, and controlling the shape of a quantum wavefunction¹⁴. Spectral resolution enhancement of the standard two-beam CARS excitation scheme by means of coherent control has recently been demonstrated^{15,16}.

In our experiment we control the population of vibrational energy levels by controlling the spectral phase of a single broadband pulse. The population of a vibrational level at energy Ω_R is proportional to $|\int d\omega E(\omega)E^*(\omega - \Omega_R)|^2$, where $E(\omega) = |E(\omega)|e^{i\Phi(\omega)}$ is the complex spectral amplitude of the applied field. Each level is thus excited by all frequency pairs separated by Ω_R . The interference between the multiple paths leading to the population of the level Ω_R is determined by the relative phase of each contribution $\Phi(\omega) - \Phi(\omega - \Omega_R)$. Constructive interference is achieved when $\Phi(\omega) = \Phi(\omega - \Omega_R)$ for all frequency components of the excitation pulse. For a transform-limited pulse (all frequency components having the same phase) this holds for all values of Ω_R , and so spectral resolution is lost. However, when the spectral phase is modulated periodically with a period Ω , constructive interference is induced for all energy levels where $\Omega_R = N\Omega$ (where N is an integer). We note that a periodic spectral phase is equivalent to splitting the pulse in time domain to several equally spaced pulses, each delayed by $\tau = 2\pi/\Omega$. A coherent population can indeed be built up only if these pulses are separated by an integer number of vibrational periods. By carefully choosing the periodic function it is possible exclusively to populate one Raman level within every frequency octave⁹. In our experiment the excited level population is probed by the excitation pulse itself. All frequency components participate, and the CARS spectrum is therefore similar to the original spectrum, upshifted by the level energy.

The measured signal consists of two components: a nonresonant contribution^{1,3}, and a resonant Raman contribution. The nonresonant contribution is due to the instantaneous electronic response of

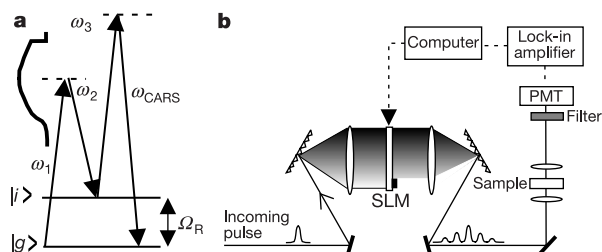


Figure 1 Graphic description of the coherent anti-Stokes Raman (CARS) process and the experimental set-up. **a**, Energy-level scheme of the single-pulse CARS process. The pulse spectrum, blocked on the high-frequency end, has a bandwidth larger than the vibrational energy level. **b**, Diagram of the experimental set-up. The pulse shaper is used to apply the desired phase function by the computer-controlled spatial light modulator (SLM), and to block the high-energy end of the spectrum. The spectral resolution of the shaper is of the order of 0.5 nm. The sample is illuminated by a microscope objective. The CARS signal at higher frequencies, blocked in the incoming pulse, is measured by the photomultiplier tube (PMT). The incoming 20-fs pulses from the Ti:sapphire oscillator are shaped into a pulse train by a periodic phase modulation.

the medium and is determined by the peak intensity of the pulse. For such short pulses it is typically over an order of magnitude stronger than the resonant contribution⁵. Suppression of this background signal is therefore crucial to the observation of any resonant signal. Pulse shaping thus serves two ends, both as a method to excite selectively a vibrational level, and to suppress the nonresonant background, which is maximized by a transform-limited pulse¹⁰, and is weaker for shaped pulses, in which the peak intensity is reduced.

We demonstrate this principle in CARS spectroscopy of several simple molecules in the liquid phase. Our experimental system consists of a mode-locked Ti:sapphire laser emitting 20-fs transform-limited pulses at 80 MHz, a programmable pulse shaper¹⁷, and a photodetector, as shown in Fig. 1b. The Fourier-transform pulse shaper is used both to apply the desired phase to each frequency component of the pulse, and to block the higher frequencies, which may overlap the CARS signal. The sample is illuminated by total bandwidth of 75 nm, equivalent to an energy span of about $1,100\text{ cm}^{-1}$. After passage through the sample, the excitation pulse is filtered out, and the CARS signal at the high-frequency end is measured. This system is suitable for CARS spectroscopy in the range of about $400\text{--}800\text{ cm}^{-1}$, typical of carbon–halogen bond stretching¹. The lower limit stems from the technical requirement to filter out the excitation pulse, and the upper limit is due to the total bandwidth.

To measure the Raman spectrum, a sinusoidal spectral phase function is applied to the pulse. The spectrum is obtained by monitoring the total CARS signal, while varying the phase-function periodicity. This, in turn, leads to a nearly sinusoidal variation in the population of any single vibrational level. For a sample with no Raman levels in the measurable range, such as methanol (see Fig. 2a), we merely observe the monotonically decreasing nonresonant signal as the number of oscillation periods is increased. When a sample with a single resonant level, such as CH_2Br_2 (resonant at 577 cm^{-1} , see Fig. 2b) is measured, the signal oscillates. The signal peaks each time the Raman level energy is commensurate with the spectral modulation period. For materials with more than one resonant level, such as $(\text{CH}_2\text{Cl})_2$ (resonant at 652 cm^{-1} and 750 cm^{-1} , see Fig. 2c) beats are observed in the signal. The Raman spectrum is obtained by Fourier transformation of the measured signal, as shown in the insets of Fig. 2. The spectral resolution of the Fourier transform operation is better than the pulse bandwidth by a factor of 40 (that is, about 30 cm^{-1}). It is limited by the maximal number of phase modulation periods on the spatial light modulator (SLM), technically determined by the number of pixels on the SLM. Thus, spectral resolution is optimized

by use of a simple sinusoidal phase function.

To achieve maximal contrast between the resonant signal and the nonresonant background, however, more complex periodic phase functions should be applied. The contrast is improved by breaking the single pulse to a train containing more pulses, each having a lower energy content, thus reducing its peak intensity. This is achieved by adding higher harmonic orders to the applied phase functions. In Fig. 3 we show how, by using a phase function containing only one additional harmonic, it is possible to attenuate the nonresonant background by nearly two orders of magnitude, while retaining the resonant component. The achieved contrast between the resonant signal and the nonresonant background is thus greatly improved. In fact, using shaped pulses we obtain a predominantly resonant signal, larger by a factor of about four than the nonresonant signal. Such phase functions should be applied when attempting to detect molecules with a known level structure.

Owing to their strong dependence on the local field, nonlinear optical processes are ideal for depth-resolved microscopy. This idea was first implemented for two-photon fluorescence microscopy¹⁸, later used for nonresonant third-harmonic generation processes^{19,20}, and has been recently extended to CARS^{21–23}. CARS microscopy has the potential for studying live biological specimens while gathering three-dimensional information on their molecular structure. Until now, however, CARS microscopes required two or three narrow-band sources. Here, we demonstrate single-pulse spectrally resolved microscopy. The sample consists of a glass capillary plate with $10\text{-}\mu\text{m}$ holes filled with CH_2Br_2 (resonant at 577 cm^{-1}). It was raster-scanned around the focused laser beam using computer-controlled piezoelectric drivers. The image in Fig. 4a was taken with a pulse shape maximizing the relative intensity of the resonant contribution, whereas the image in Fig. 4b was taken with a pulse shape minimizing this intensity. The predominantly resonant signal from the filled holes in Fig. 4a is larger by a factor of four than in Fig. 4b, and the signal from the glass is very similar. Figure 4c shows the difference between Fig. 4a and b, depicting the signal from the 577 cm^{-1} vibrational level of CH_2Br_2 . This image appears inverted relative to that obtained using a transform-limited pulse (Fig. 4d), where the glass, having a larger nonresonant signal, appears brighter. The image in Fig. 4c demonstrates the ability of the microscope to spectrally resolve the Raman resonant contribution of a single vibrational level. In a practical system, the difference image could be directly measured by lock-in detection, alternating the phase masks of Fig. 4a and b at a high frequency.

For materials having more than one vibrational level it is possible

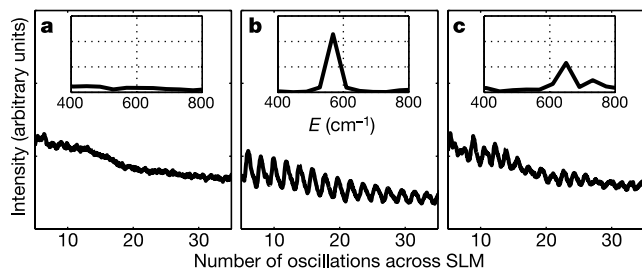


Figure 2 Single-pulse CARS spectroscopy of various organic molecules in the liquid phase. The CARS signal intensity is shown as a function of the number of periods of the sinusoidal phase across the SLM for: **a**, methanol; **b**, CH_2Br_2 ; **c**, $(\text{CH}_2\text{Cl})_2$. The insets show the respective measured Raman spectra, derived by Fourier transformation of the intensity curves. In methanol (**a**), which has no resonance at the measured frequency range, only a monotonically decreasing nonresonant signal is observed. In CH_2Br_2 (**b**), which has a single resonance at $\Omega_R = 577\text{ cm}^{-1}$, the signal peaks periodically, whenever the modulation period is an integer fraction of Ω_R . The Fourier transform operation retrieves the resonance with a resolution that is inversely proportional to the number of observed periods. In $(\text{CH}_2\text{Cl})_2$ (**c**), beating between the 652 cm^{-1} and the 750 cm^{-1} levels is observed.

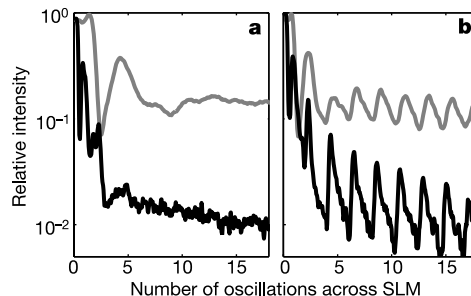


Figure 3 Demonstration of the nonresonant background suppression. The CARS signal intensity is shown as a function of the number of phase oscillation periods across the SLM for both a sinusoidal phase function (grey line) and a phase function containing an additional harmonic component (solid line). The CARS signal is plotted relative to that obtained by a transform-limited pulse (0 phase function oscillations). Results are for: **a**, methanol (nonresonant component only); **b**, CH_3I (a single resonance at 523 cm^{-1}). Suppression of the nonresonant background by nearly two orders of magnitude is achieved (**a**), while the resonant component is almost completely restored. To achieve adequate suppression of the nonresonant background at least several oscillation periods of the phase function across the pulse spectrum are necessary.

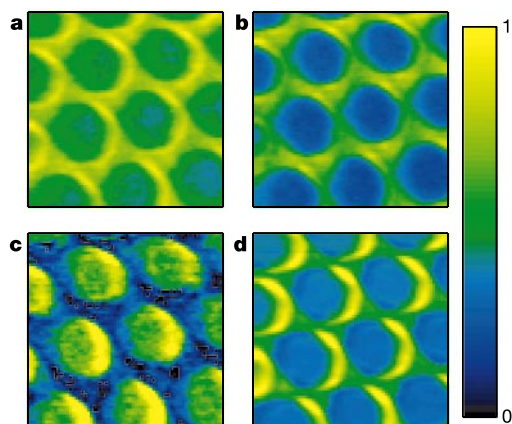


Figure 4 Depth-resolved single-pulse CARS images of a glass capillary plate with 10- μm holes filled with CH_2Br_2 (resonant at 577 cm^{-1}). The average laser power was about 5 mW. **a**, Image obtained using a shaped pulse (similar to the high-contrast one in Fig. 3) maximizing the resonant contribution of the CH_2Br_2 577 cm^{-1} level, while reducing the nonresonant contribution. **b**, Image obtained using a shaped pulse minimizing the resonant contribution of the CH_2Br_2 577 cm^{-1} level. **c**, Difference image, obtained by subtraction of **a** and **b**. This procedure retains only the 577 cm^{-1} resonant component. The image appears inverted relative to an image obtained using a transform-limited pulse (**d**), because a resonant signal appears only in the hole regions containing CH_2Br_2 whereas the nonresonant signal is stronger in the glass. The $45\text{ }\mu\text{m} \times 45\text{ }\mu\text{m}$ images were obtained using an objective of numerical aperture 0.65, at a rate of 30 ms per pixel. The images are plotted using an arbitrary scale, although in **a** and **b** the same scale is used.

to improve the detection selectivity by tailoring shaped pulses to induce constructive quantum interference of these levels. A simple example is the shape used to obtain peaks in the beat signal of Fig. 2c, where the resonant contributions of two levels of $(\text{CH}_2\text{Cl})_2$ constructively interfere.

Our experiment shows how single-pulse vibrational microspectroscopy can be readily achieved by using coherent control techniques. By tailoring the spectral phase of an ultrashort pulse, we are able to control the interference between the various spectral components of the pulse, leading to selective population of given vibrational energy levels. By applying this principle, we have demonstrated a robust method for high-resolution spectroscopy, nearly two orders of magnitude greater than the spectral bandwidth, in the vibrational energy range $400\text{--}800\text{ cm}^{-1}$. This method is easily extendable to the fingerprint region ($1,000\text{--}1,500\text{ cm}^{-1}$) by using slightly shorter pulses, available in commercial systems today. Single-pulse CARS is particularly suitable for nonlinear microscopy. In contrast to conventional nonlinear CARS microscopy, which requires two synchronized tunable sources, single-pulse CARS requires only a single ultrashort laser that does not need to be tuned. The spectral measurements are performed with an electronically controlled SLM. Moreover, it can be easily combined with other nonlinear microscopic methods such as multiphoton fluorescence and third-harmonic generation using the same microscope set-up. The concept of performing a nonlinear optical interaction with matter in a single coherently controlled pulse offers a promising alternative to standard multiple-pulse nonlinear systems in use today. We believe that combined with present-day sources, covering most of the optical spectrum in a single pulse, this concept will have a significant impact on nonlinear spectroscopy and microscopy applications. □

Received 8 January; accepted 19 June 2002; doi:10.1038/nature00933.

1. Kiefer, W. in *Infrared and Raman Spectroscopy* (ed. Schrader, B.) 162–188 (VCH, Weinheim, 1995).
2. Yan, Y. X., Cheng, L. T. & Nelson, K. A. in *Advances in Spectroscopy* (eds Clark, R. J. H. & Hester, R. E.) Vol. 16, 299–355 (Wiley, Chichester, 1988).
3. Levenson, M. D. *Introduction to Nonlinear Laser Spectroscopy* (Academic, New York, 1982).
4. Leonhardt, R., Holzapfel, W., Zinth, W. & Kaiser, W. Terahertz quantum beats in molecular liquids.

Chem. Phys. Lett. **133**, 373–377 (1986).

5. Cheng, J., Volkmer, A., Book, L. D. & Xie, X. S. An epi-detected coherent anti-Stokes Raman scattering (E-CARS) microscope with high spectral resolution and high sensitivity. *J. Phys. Chem. B* **105**, 1277–1280 (2001).
6. Tannor, D. J. & Rice, S. A. Control of selectivity of chemical reaction via control of wavepacket evolution. *J. Chem. Phys.* **83**, 5013–5018 (1985).
7. Shapiro, M. & Brumer, P. Laser control of product quantum state populations in unimolecular reactions. *J. Chem. Phys.* **84**, 4103–4104 (1986).
8. Warren, W. S., Rabitz, H. & Dahleh, M. Coherent control of quantum dynamics: the dream is alive. *Science* **259**, 1581–1589 (1993).
9. Weiner, A. M., Leaird, D. E., Wiederrecht, G. P. & Nelson, K. A. Femtosecond pulse sequences used for optical manipulation of molecular motion. *Science* **247**, 1317–1319 (1990).
10. Meshulach, D. & Silberberg, Y. Coherent quantum control of two-photon transitions by a femtosecond laser pulse. *Nature* **396**, 239–242 (1998).
11. Judson, S. R. & Rabitz, H. Teaching lasers to control molecules. *Phys. Rev. Lett.* **68**, 1500–1503 (1992).
12. Assion, A. *et al.* Control of chemical reaction by feedback-optimized phase-shaped femtosecond laser pulses. *Science* **282**, 919–922 (1998).
13. Bartels, R. *et al.* Shaped pulse optimization of coherent emission of high-harmonic soft X-rays. *Nature* **406**, 164–166 (2000).
14. Weinacht, T. C., Ahn, J. & Bucksbaum, P. H. Controlling the shape of a quantum wavefunction. *Nature* **397**, 233–235 (1999).
15. Oron, D., Dudovich, N., Yelin, D. & Silberberg, Y. Quantum control of coherent anti-Stokes Raman processes. *Phys. Rev. A* **65**, 043408-1–043408-4 (2002).
16. Oron, D., Dudovich, N., Yelin, D. & Silberberg, Y. Narrow band coherent anti-Stokes Raman signals from broadband pulses. *Phys. Rev. Lett.* **88**, 063004-1–063004-4 (2002).
17. Weiner, A. M. Femtosecond pulse shaping using spatial light modulators. *Rev. Sci. Instrum.* **71**, 1929–1960 (2000).
18. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
19. Barad, Y., Eisenberg, H., Horowitz, M. & Silberberg, Y. Nonlinear scanning laser microscopy by third harmonic generation. *Appl. Phys. Lett.* **70**, 922–924 (1997).
20. Muller, M., Squier, J., Wilson, K. R. & Brakenhoff, G. J. 3D microscopy of transparent objects using third harmonic generation. *J. Microsc.* **191**, 266–274 (1998).
21. Zumbusch, A., Holtom, G. R. & Xie, X. S. Three-dimensional vibrational imaging by coherent anti-Stokes Raman scattering. *Phys. Rev. Lett.* **82**, 4142–4145 (1999).
22. Potma, E. O., de Boeij, W. P., van Haastert, P. J. M. & Wiersma, D. A. Real-time visualization of intracellular hydrodynamics in single living cells. *Proc. Natl Acad. Sci. USA* **98**, 1577–1582 (2001).
23. Muller, M., Squier, J., de Lange, C. A. & Brakenhoff, G. J. CARS microscopy with folded BoxCARS phasematching. *J. Microsc.* **197**, 150–158 (2000).

Acknowledgements

We thank D. Mandelik for his aid in preparing the microscope samples. Financial support by the Israeli Science Foundation and by the Bundesministerium für Bildung und Forschung is gratefully acknowledged.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.S. (e-mail: yaron.silberberg@weizmann.ac.il).

A large-cavity zeolite with wide pore windows and potential as an oil refining catalyst

Avelino Corma*, María J. Díaz-Cabañas*, Joaquín Martínez-Triguero*, Fernando Rey* & Jordi Rius†

* Instituto de Tecnología Química, UPV-CSIC, Universidad Politécnica de Valencia, Avda. de los Naranjos s/n, 46022 Valencia, Spain

† Institut de Ciència de Materials de Barcelona (CSIC), Campus de la UAB, 08193-Bellaterra, Catalunya, Spain

Crude oil is an important feedstock for the petrochemical industry and the dominant energy source driving the world economy, but known oil reserves will cover demand for no more than 50 years at the current rate of consumption¹. This situation calls for more efficient strategies for converting crude oil into fuel and petrochemical products. At present, more than 40% of oil conversion is achieved using catalysts based on

faujasite; this zeolite requires extensive post-synthesis treatment to produce an ultrastable form^{2,3}, and has a large cavity accessible through four 0.74-nm-wide windows and thus limits the access of oil molecules to the catalytically active sites. The use of zeolites with better accessibility to their active sites should result in improved catalyst efficiency. To date, two zeolites with effective pore diameters exceeding that of faujasite have been reported^{4,5}, but their one-dimensional pore topology excludes use in oil refining. Similarly, zeolites with large pores and a three-dimensional pore topology have been reported^{6–8}, but in all these materials the pore openings are smaller than in faujasite. Here we report the synthesis of ITQ-21, a zeolite with a three-dimensional pore network containing 1.18-nm-wide cavities, each of which is accessible through six circular and 0.74-nm-wide windows. As expected for a zeolite with this structure, ITQ-21 exhibits high catalytic activity and selectivity for valuable products in preliminary oil refining tests.

Zeolite Y or faujasite⁹ consists of large cavities interconnected by four large windows. It can only be synthesized with high aluminium content, and therefore has to be submitted to dealumination and stabilization processes before its use in fluid catalytic cracking and hydrocracking of heavy oil fractions^{2,3}. New zeolites with large pores, three-dimensional pore topologies and adequate Si/Al ratios could form the basis of improved catalysts for these cracking processes and thus improve oil conversion, but such materials have proved difficult to prepare. We approached this problem by combining the use of large organic cations as structure-directing agents with the use of germanium, which is known to direct zeolite formation towards structures with double four-membered rings (D4MR)^{8,10,11}. We anticipated that this combination could lead to zeolites with large void spaces and/or pores (that is, low framework density).

Zeolite A⁹ is the zeolite with the lowest framework density containing D4MR units. Like zeolite Y, it consists of sodalite cages, but they are connected by D4MR units to create cavities; and whereas the cavities in zeolite Y are interconnected through four windows, the cavities in zeolite A each contain six such windows. The larger number of connecting windows thus favours diffusion through zeolite A over diffusion through zeolite Y, yet the smaller window dimension (0.45 nm) in the former limits the practical utility of this material. Overall, it thus appears that a zeolite structure with cavities at least as spacious as those of zeolite Y but interconnected through six 12-membered-ring windows of 0.74 nm might be interesting for catalytic applications. The larger diffusivity of molecules through the structure of such a material would allow

efficient gas-oil cracking, and produce a lower extent of olefin saturation in the gasoline product. The larger diffusivity should also facilitate re-cracking of the olefins versus hydrogen transfer, thus yielding more of the industrially preferred propylene.

We synthesized the new zeolite ITQ-21, which possesses a topology similar to the hypothetical structure described above, by using N(16)-methylsparteinium cation (MSPT) as a bulky and rigid organic structure-directing agent, in combination with fluoride anions and germanium^{8,12}.

The powder X-ray diffraction (XRD) pattern of the as-synthesized sample with a Si/Ge ratio of 1.91 is shown in Fig. 1. From this, the crystal structure of ITQ-21 was determined, assuming space group symmetry *Fm-3c* with $a = 27.698 \text{ \AA}$ (see Methods for details). The ITQ-21 structure can be best described by considering a cluster formed by a $[4^6 6^{12}]$ cage with a four-membered ring (4MR) inside (Fig. 2 inset), linked to six neighbouring similar clusters thereby forming D4MR cages (Fig. 2). This yields 8 clusters and 24 D4MR cages in the unit cell, which has cubic cell parameters. The inner 4MRs of the clusters are thus most probably randomly oriented normal to the cell axes, although crystal twinning with ordered domains and a tetragonal structure would also yield cubic cell parameters.

The channel system in ITQ-21 is three-dimensional, with circular openings comprising 12-membered rings. The three linear channels intersect to produce large inner cavities, with a nearly spherical shape about 1.18 nm in diameter. (See Methods for details on the structure solution; see Supplementary Information Table 1 for fractional atomic coordinates for the 'as synthesized' ITQ-21.) The small amorphous content present in the material (4.3%) was estimated from the refined Si/Ge ratio in the crystal structure (1.71) and by assuming that no Ge is in the amorphous phase (from our experience, when germanium is not incorporated in the structure, it appears as quartz-like GeO_2 that is easily detected by XRD). As the crystal structure of ITQ-21 was solved on the as-prepared material, the structure contains N-methylsparteinium cations. According to the corrected chemical analysis, there are 22–24 SDA cations and 15.5 F^- in the unit cell. XRD indicates that the centres of the D4MR cages are fully occupied either by F^- or $\text{H}_2\text{O}/\text{OH}^-$. It is difficult to differentiate between these species by XRD owing to their similar scattering powers, but ^{19}F magic-angle spinning (MAS) NMR

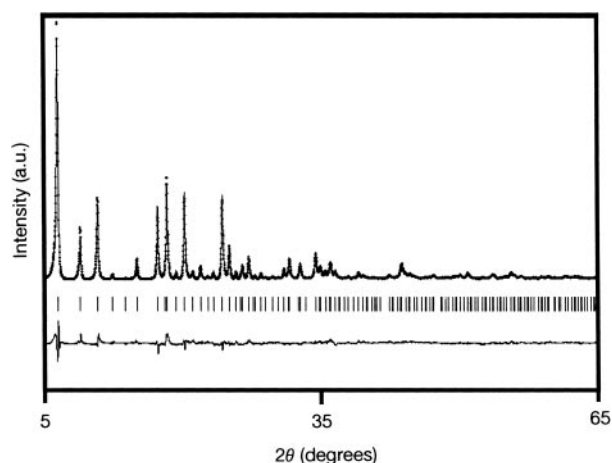


Figure 1 Rietveld refinement of ITQ-21. The experimental (crosses) and calculated (lines) XRD patterns are shown, as well as the difference profile (bottom). The short tick marks below the pattern give the positions of the Bragg reflections.

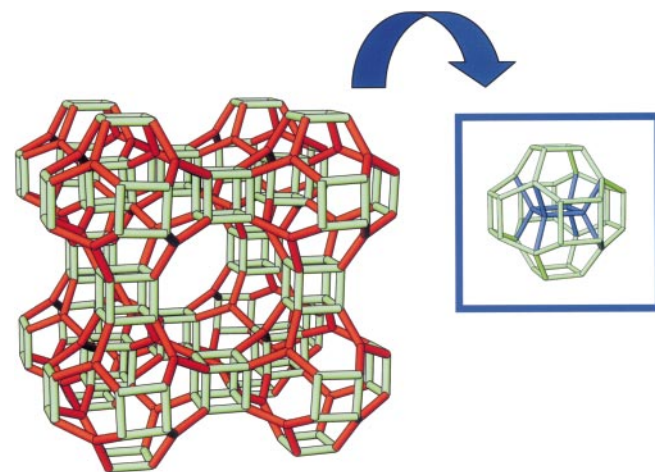


Figure 2 Perspective view of the unit cell of ITQ-21. It shows the outer T atoms of the clusters—that is, cage $[4^6 6^{12}]$ —interconnected via double four-membered rings (D4MR, shown as green lines; linkages between neighbouring D4MR cages are shown in red). Each tetrahedral intersection represents one T site (either occupied by Si or Ge) bridged by O atoms. Inset, the detailed structure of the complete cluster showing the four-membered ring inside (shown as blue lines). This ring is assumed to be randomly oriented normal to the three cell axes.

spectroscopy of the as-synthesized ITQ-21 provides clear evidence for the presence of F^- anions in the D4MR cages—a single band at about -8 p.p.m. (refs 8, 13). As the chemical analysis gives approximately $15.5 F^-$ per unit cell, we conclude that in this sample nearly two-thirds of the 24 DMR units per unit cell are occupied by F^- and the rest must be occupied by H_2O/OH^- . The N-methylsparteinium cation is particularly difficult to handle in the refinement owing to the lack of internal symmetry, but its averaged electron density could be modelled with two peaks, each possessing a scattering power equal to half the number of electrons in the molecule. From the location of these peaks, we see that the extra-framework space is filled efficiently.

ITQ21 contains three symmetrically independent T sites. Their atomic compositions are 45(2)% Ge for T1, 31(2)% Ge for T2 and 0(3)% Ge for T3. As T1 is located at the D4MR units, this indicates a preferential occupation of Ge atoms in the D4MR units. A similar Ge distribution has been observed in pure polymorph C of zeolite beta⁸.

One sample of aluminium-containing ITQ-21 ((Si + Ge)/Al = 25) was synthesized (sample B in Methods); the XRD patterns of the as-made and the calcined Al-ITQ-21 sample are available (see Supplementary Information Fig. 1). The micropore volume of this material ($0.24 \text{ cm}^3 \text{ g}^{-1}$) indicates that it is highly crystalline. Therefore, the broadening of the X-ray reflections upon Al incorporation should be attributed to the observed decrease in the average crystal size of the sample.

The ^{27}Al MAS NMR spectrum of the sample calcined at 823 K and rehydrated presents (data not shown) an intense peak at 54 p.p.m. attributed to tetrahedrally coordinated Al, which is generally taken as a proof of the isomorphic substitution of Si by Al (ref. 14). A much smaller peak at 0 p.p.m., associated with octahedrally coordinated and probably extra-framework Al, indicates that on calcination most of the Al remains in framework positions. This interpretation is further supported by the infrared spectra of the OH region (Fig. 3a), and by stepwise pyridine desorption (Fig. 3b) of the Al-ITQ-21 sample. Indeed, the calcined Al-ITQ-21 shows an absorption at $3,620 \text{ cm}^{-1}$ attributed to acid bridging hydroxyl groups that disappears when adsorbing pyridine (Fig. 3a). Meanwhile, pyridinium cations are formed, as deduced from the characteristic band at $1,545 \text{ cm}^{-1}$. The band at $1,450 \text{ cm}^{-1}$, assigned to pyridine interacting with Lewis acid sites, would indicate the presence of extra-framework aluminium species in the sample. The retention of pyridinium species upon desorption at 623 K

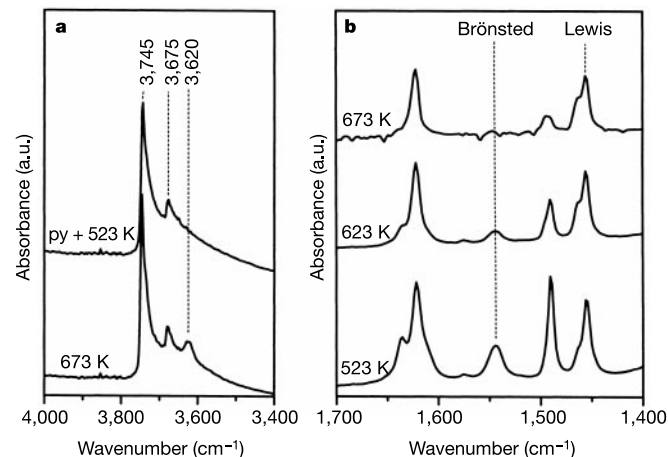


Figure 3 Infrared spectra of Al-ITQ-21. **a**, The OH stretching region after outgassing of the sample at 673 K, pyridine adsorption at room temperature and subsequent desorption at 523 K. **b**, The C=C and C=N stretching region of the adsorbed pyridine after desorption at 523, 623 and 673 K.

Table 1 Gas-oil cracking activity and selectivities

Sample	Zeolite				
	ITQ-21	Beta	USY-1	Steamed ITQ-21	Steamed USY-1
Gas-oil cracking conversion*, wt%	72.5	53.9	68.0	41.5	36.3
Gasoline yield†, wt%	30.8	27.8	39.2	29.6	31.3
Diesel yield†, wt%	12.0	9.2	13.9	12.9	15.1
Gases yield†, wt%	24.23	28.4	14.7	16.6	12.6
Coke yield†, wt%	3.0	4.7	2.2	3.0	3.1
Propylene yield†, wt%	6.0	5.8	3.5	4.4	3.3
Olefins in gasoline, wt%	11.6	14.8	14.0	15.9	22.6
RON	89.9	88.2	86.9	86.2	85.9
MON	85.0	83.9	82.4	81.9	81.4

* Catalyst contact time: catalyst/oil ~0.35, time on stream 60 s, reaction temperature 773 K.

† Yields of calcined and steamed samples are calculated at 70 and 62 wt% conversion, respectively. Conversion is defined as the sum of gasoline, diesel, gases and coke.

under vacuum (Fig. 3b) clearly shows the presence of strong Brønsted acid sites in this sample.

The cracking activity of this sample was studied with an industrial feed—Arabian light vacuum gas oil (gas-oil characteristics are available; see Supplementary Information Table 2). Table 1 compares the cracking performance obtained with Al-ITQ-21 and two commercial zeolites, USY-1 and beta, which also have a three-dimensional pore topology and large pores. (Physicochemical characteristics of the three catalyst samples are given in Supplementary Information Table 3) The results show that the catalytic activity of ITQ-21 is higher than that of USY-1, and much higher than that of zeolite beta, even though the average crystal size of the latter is similar to that of the ITQ-21 sample used. On the other hand, ITQ-21 produces a higher propylene yield and a higher propylene concentration in the gas fraction than USY-1 zeolite, and slightly larger than beta even if this produces more gases than ITQ-21. The gasoline analysis shows that the olefin concentration is the lowest with ITQ-21, while the octane number is the highest.

The results of the catalytic tests thus confirm our hypothesis that a material with a structure similar to that of ITQ-21 should exhibit similar gas-oil cracking activity to that of USY zeolite while producing less hydrogen transfer and a higher propylene yield, owing to the combination of pore topology, crystal size and acidity. These effects are still observed when the cracking behaviour of ITQ-21 is compared to that of a catalyst blend formed by a 1.0:0.2 weight ratio of USY and ZSM-5 (Si/Al = 40) zeolites, which corresponds to the actual technology used to maximize propylene yields. ITQ-21 produces a 6.3% yield of propylene while the blend formed by USY and ZSM-5 produces 5.7%, at 70% of total conversion.

ITQ-21 and USY samples have also been tested for cracking methylcyclohexane, which provides a test reaction for assessing the extent of hydrogen transfer^{15–17}. The results obtained show that selectivity to toluene and benzene is lower with the ITQ-21 (5.04 wt% of toluene and 0.23 wt% of benzene) than with the USY sample (6.25 wt% of toluene and 0.47 wt% of benzene), and also the isobutene to isobutane ratio is higher with the ITQ-21 sample (0.034 versus 0.018) at 70 wt% of total conversion. These results agree with our proposal on the lower hydrogen transfer activity of ITQ-21.

ITQ-21 is stable on steaming, and a surface area close to $300 \text{ m}^2 \text{ g}^{-1}$ remains after calcination at 993 K in 100% steam for 5 hours (Table 1). When the steamed ITQ-21 sample is compared with a steamed USY-1 in cracking, the former is more active and produces more propylene in the gas fraction while giving a less olefinic gasoline with higher RON and MON (the research and motor octane numbers, respectively). These preliminary laboratory results indicate that ITQ-21 is promising for catalytic cracking applications. □

Methods

Synthesis of ITQ-21 samples

We used precursor gels A (ITQ-21) and B (Al-ITQ-21) with the following compositions. A, 0.33 GeO₂: 0.67 SiO₂: 0.50 MSPTOH: 0.50 HF: 20 H₂O. B, 0.09 GeO₂: 0.91 SiO₂: 0.02 Al₂O₃: 0.50 MSPTOH: 0.50 HF: 3 H₂O. They were prepared by dissolving germanium oxide in the N-methylsperminium hydroxide (MSPTOH) solution. Then, tetraethylorthosilicate and aluminium isopropoxide (for the aluminium-containing sample) were hydrolysed in that solution, and the mixture was mechanically stirred at room temperature until complete elimination of the alcohols had occurred, and the required amount of water until the final composition was reached. Finally, HF was added and the mixture was homogenized.

When the gels were heated at 448 K for five days under conditions of continuous stirring in Teflon-lined stainless-steel autoclaves, the zeolites crystallized. The solids were recovered by filtration, washed with distilled water and dried at 100 °C. Chemical analyses of the ITQ-21 samples presented here are given as Supplementary Information Table 4.

Structure solution and refinement

Powder XRD data were collected at room temperature with a diffractometer (Philips X'Pert) with Bragg-Brentano geometry, and a graphite secondary monochromator. Intensity data were obtained with a fixed divergence slit (1/8°), a maximum illumination length of 10 mm and a goniometer arm of 240 mm, and CuKα₁₂ radiation (λ = 1.5406, 1.5441 Å). Tube voltage and current were respectively 45 kV and 40 mA. Step scan size and time: 0.02° (2θ) and 9 s. The profile range (2θ) goes from 5.00° to 70.00°. The pattern was indexed according to a cubic F unit cell with a = 27.698 Å, and the analysis of the observed systematic absences led to space group *Fm-3c*.

The crystal structure was solved using XLENS¹⁸ with space group *Fm-3c*, no. 226, that only contains three symmetry independent T sites (T1, T2 and T3). A Rietveld refinement in this cubic symmetry yielded the atomic compositions of sites T1, T2 and T3, and also indicated that the sites containing F[−] or H₂O/OH[−] are fully occupied. In space group *Fm-3c*, the cluster shown in Fig. 2 inset is placed at (1/4, 1/4, 1/4), and this site has 43 point group symmetry, so that the Si atoms forming the inner 4MR are averaged over 12 T3 sites. The Rietveld refinement was performed with program LSP7¹⁹ yielding the atomic coordinates listed in Supplementary Information Table 1. Number of contributing reflections is 224; number of geometric restraints of type d(T–O) is 12 (for T1: 1.69(3) Å; T2: 1.66(3); T3: 1.61(3)). Interatomic distances of type O–T–O were also used as restraints with a standard deviation of 0.06 Å. Number of structural parameters: 23. Number of profile parameters: 7, including zero shift (0.029° 2θ) with visually estimated background. Profile function: Pearson-VII. Rietveld overall thermal coefficient: B = 2.9 Å². The residuals of the refinement were: R_{wp} = 0.083; R_p = 0.063, R₁ = 0.127. The good fit between observed and calculated patterns can be seen in Fig. 1.

Also, the possibility that all clusters are equally oriented has been checked by assuming the tetragonal space group *I4/mcm* no. 140 (a_T ≈ a/√2, c_T ≈ a). In spite of the much larger number of parameters involved (50), the improvement of the residuals is minimal. In addition, if the topology were tetragonal, the structure along the c axis would be significantly different from that along the a axis, and this should be evident in the cell parameters. However, the refined tetragonal parameters a_T = 19.5896(10) and c_T = 27.6946(33) Å correspond, within the statistical error, to the cubic cell. Nevertheless, a tetragonal structure cannot be completely ruled out, as the presence of twinned crystals with ordered domains would also result in cubic cell parameters, despite the tetragonal structure. High-resolution electron microscopy investigations might help to clarify this aspect of the ITQ-21 structure.

Catalytic tests

The catalytic experiments were performed in a Microactivity Test Unit (ASTM D-3907). Product determination, including a detailed analysis of the gasoline fraction—PIONA (gas chromatographic analysis of paraffins, isoparaffins, olefins, naphthenes and aromatic hydrocarbon types in gasoline), RON and MON numbers—were performed as described²⁰. Zeolites were pelletized, crushed and sieved, and 0.50 g of the 0.59–0.84-mm fraction was taken and diluted in 2.50 g of inert silica. In the case of the USY/ZSM-5 mixture, they were placed in consecutive beds as described previously²¹. The catalyst/oil ratio was calculated by dividing the amount of zeolite by the amount of gas oil fed.

Received 2 October 2001; accepted 18 June 2002; doi:10.1038/nature00924.

1. Maricly, C. Evolution of refining and petrochemicals. What is the place of zeolites? *Stud. Surf. Sci. Catal.* **135**, 37–60 (2001).
2. Biswas, J. & Maxwell, I. E. Recent process-related and catalyst-related developments in fluid catalytic cracking. *Appl. Catal.* **63**, 197–258 (1990).
3. Martino, G. Catalysis for oil refining and petrochemistry, recent developments and future trends. *Stud. Surf. Sci. Catal.* **A 130**, 83–103 (2000).
4. Freyhardt, C. C., Tsapatsis, M., Lobo, R. F., Balkus, K. J. Jr & Davis, M. E. A high-silica zeolite with a 14-tetrahedral-atom pore opening. *Nature* **381**, 295–298 (1996).
5. Yoshikawa, M. *et al.* Synthesis, characterization, and structure solution of CIT-5, a new, high-silica, extra-large-pore molecular sieve. *J. Phys. Chem. B* **102**, 7139–7147 (1998).
6. Treacy, M. M. J. & Newsam, J. M. Two new three-dimensional twelve-ring zeolite frameworks of which zeolite beta is a disordered intergrowth. *Nature* **332**, 249–251 (1988).
7. Cambor, M. A. *et al.* High silica zeolites with three-dimensional systems of large pore channels. *Micropor. Mesopor. Mater.* **48**, 11–22 (2001).
8. Corma, A., Navarro, M. T., Rey, F., Rius, J. & Valencia, S. Pure polymorph C of zeolite beta synthesized by using framework isomorphous substitution as a structure-directing mechanism. *Angew. Chem. Int. Edn Engl.* **40**, 2277–2280 (2001).
9. Baerlocher, Ch., Meier, W. M. & Olson, D. H. *Atlas of Zeolite Framework Types* 5th edn (Elsevier, Amsterdam, 2000).

10. O'Keeffe, M. & Yaghi, O. M. Germanate zeolites: contrasting the behaviour of germanate and silicate structures built from cubic T₈O₂₀ units (T = Ge or Si). *Chem. Eur. J.* **5**, 2796–2801 (1999).
11. Corma, A., Navarro, M. T., Rey, F. & Valencia, S. Synthesis of pure polymorph C of beta zeolite in a fluoride-free system. *Chem. Commun.* **16**, 1486–1487 (2001).
12. Corma, A., Diaz-Cabanas, M. J. & Fornes, V. Synthesis, characterization, and catalytic activity of a large-pore tridirectional zeolite, H-ITQ-7. *Angew. Chem. Int. Edn Engl.* **39**, 2346–2349 (2000).
13. Blasco, T. *et al.* Preferential location of Ge in the double four-membered ring units of ITQ-7 zeolite. *J. Phys. Chem. B* **106**, 2634–2642 (2002).
14. Klinowski, J. Nuclear magnetic resonance of zeolites. *Prog. Nucl. Magn. Reson. Spectrosc.* **16**, 237–309 (1984).
15. Cerqueira, H. S., Mihindou-Koumba, P. C., Magnoux, P. & Guisnet, M. Methylcyclohexane transformation over HFAU, HBEA, and HMFI zeolites: I. Reaction scheme and mechanisms. *Ind. Eng. Chem. Res.* **40**, 1032–1041 (2001).
16. Corma, A. *et al.* Methylcyclohexane and meoxythylcyclohexene cracking over zeolite Y catalysts. *Appl. Catal.* **67**, 307–324 (1991).
17. Lacombe, S., Patriceon, A. & Benazzi, E. n-heptane hydroconversion and methylcyclohexane cracking as model reactions to investigate the pore topology of NU-88 zeolite. *Stud. Surf. Sci. Catal.* **135**, 4272–4279 (2001).
18. Rius, J. XLENS, a direct methods program based on the modulus sum function: Its application to powder data. *Powder Diff.* **14**, 267–273 (1999).
19. Rius, J. Program LSP7. A program for restrained Rietveld refinement (Institut de Ciència de Materials de Barcelona-CSIC, 1999).
20. Corma, A., Martínez-Triguero, J. & Martínez, C. The use of ITQ-7 as a FCC zeolitic additive. *J. Catal.* **197**, 151–159 (2001).
21. Corma, A. & Martínez-Triguero, J. The use of MCM-22 as a cracking zeolitic additive for FCC. *J. Catal.* **165**, 102–120 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. L. Jordá for partial analyses for the structure determination and for discussions. J.R. thanks the "D. Gral de Investigación" of the MCYT for support. This work was supported by the CICYT.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.C. (e-mail: acorma@itq.upv.es).

Lamellar magnetism in the haematite–ilmenite series as an explanation for strong remanent magnetization

Peter Robinson^{*†}, Richard J. Harrison[‡], Suzanne A. McEnroe^{*†} & Robert B. Hargraves[§]

^{*} Geological Survey of Norway, N-7491, Trondheim, Norway

[†] Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003, USA

[‡] Department of Earth Sciences, University of Cambridge, Cambridge, UK

[§] Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA

Magnetic anomalies associated with slowly cooled igneous and metamorphic rocks are commonly attributed to the presence of the mineral magnetite. Although the intermediate members of the ilmenite–haematite mineral series can also carry a strong ferrimagnetic remanence¹, it is preserved only in rapidly cooled volcanic rocks, where formation of intergrowths of weakly magnetic haematite and paramagnetic ilmenite is suppressed. But the occurrence of unusually large and stable magnetic remanence in rocks containing such intergrowths has been known for decades^{2–5}, and has recently been the subject of intense investigation^{6–10}. These unmixed oxide phases have been shown

to contain pervasive exsolution lamellae with thickness from 100 μm down to about 1 nm (one unit cell). These rocks, many of which contain only a few per cent of such oxides, show natural remanent magnetizations up to 30 A m^{-1} —too strong to be explained even by pure haematite in an unsaturated state^{11,12}. Here we propose a new ferrimagnetic substructure created by ferrous–ferric ‘contact layers’ that reduce charge imbalance along lamellar contacts between antiferromagnetic haematite and paramagnetic ilmenite. We estimate that such a lamellar magnetic material can have a saturation magnetization up to 55 kA m^{-1} —22 times stronger than pure haematite—while retaining the high coercivity and thermal properties of single-domain haematite.

Slowly cooled rocks from New York⁶, Sweden⁷ and Norway^{8–10}, containing finely exsolved members of the haematite–ilmenite (Fe_2O_3 – FeTiO_3) series, have strong and extremely stable remanent magnetization, suggesting an explanation for some magnetic anomalies in the deep Earth and on planetary bodies that no longer retain a magnetic field, such as Mars. Common to these rocks are grains of ilmenite–haematite containing multiple generations of exsolution lamellae, with thickness ranging down to unit-cell scale (1–2 nm; Fig. 1). These were magnetized about one billion years ago, and are magnetically extremely ‘hard’ (many do not demagnetize in alternating fields of 0.1 T, the typical maximum alternating field used in palaeomagnetic laboratories). Thermal experiments show that demagnetization occurs between 530 °C and 650 °C, comparable to, or higher than, near-endmember magnetite, and consistent with slightly titanian haematite. Haemo-ilmenite samples without coexisting magnetite have natural remanent magnetization to magnetic saturation ratios of only $\sim 2.5\%$, showing that the haematite component is in an unsaturated state. These observations led to a theoretically unexplained but firm conviction that the magnetic remanence of these materials relates to the abundance of exsolution lamellae produced during slow cooling, and is strongly coupled to

the antiferromagnetic sublattice of the haematite lamellae^{7,8,13}. Despite extensive study, the magnetic properties of these rocks could not be reconciled with their magnetic petrology using conventional rock-magnetic theory, and required a new departure.

The crystal structures of ilmenite and haematite consist of stacked cation layers^{14,15}. The magnetic moments on alternating layers (referred to as α and β) align antiparallel to each other. The net magnetic moment of the structure is given by the difference between the α and β sublattice magnetizations, which in turn is determined by the partitioning of Fe^{2+} , Fe^{3+} and Ti between the layers. In pure haematite, both α and β are occupied by Fe^{3+} , resulting in an antiferromagnetic structure with zero net moment (in reality, the antiparallel alignment is slightly canted, leading to a small parasitic moment perpendicular to the alignment of the spins). In haematite-rich titanohaematite, Fe^{2+} , Fe^{3+} and Ti are randomly distributed over α and β , again leading to an antiferromagnetic structure with zero net moment. In ilmenite-rich ferri-ilmenite, layers containing mixtures of Fe^{2+} and Fe^{3+} alternate with layers containing mixtures of Ti and Fe^{3+} . The unequal distribution of Fe leads to a net ferrimagnetic moment. Because of the high Ti content, however,

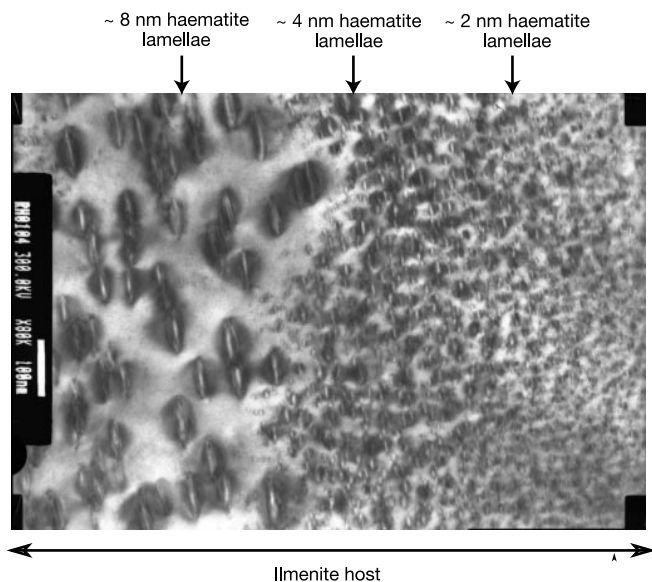


Figure 1 Transmission electron microscope image of haematite exsolution lamellae parallel to (001) in an ilmenite host from South Rogaland, Norway. Three thickness classes (from left to right) of about 8 nm, 4 nm and 2 nm are shown, interpreted as the result of three generations of lamellar nucleation and growth with slowly falling temperature. White scale bar, 100 nm. Dark fringes around lamellae indicate lattice strain in the host adjacent to coherent lamellar boundaries, postulated to increase magnetic coercivity¹⁸.

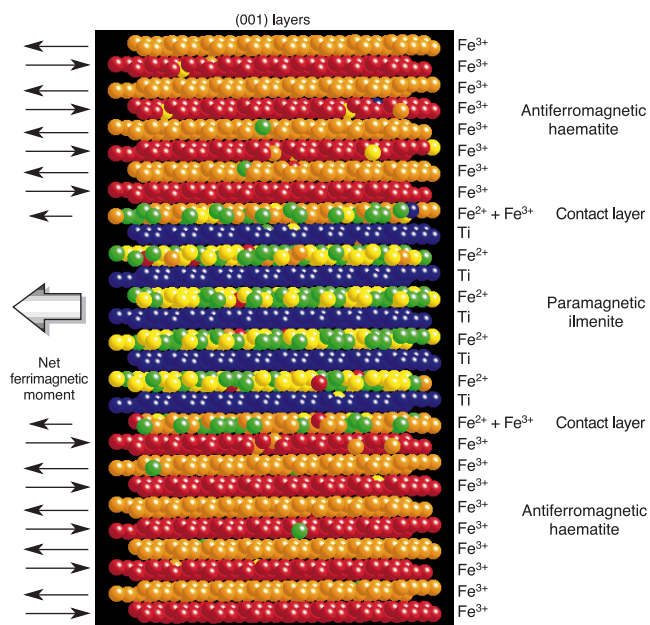


Figure 2 Part of a Monte Carlo simulation of cation ordering, magnetic ordering, and exsolution in the ilmenite–haematite system. 28 layers of a 40-layer cell are shown. Simulation was performed at 500 K and bulk composition 25 mol% ilmenite (Ilm_{25}). During the simulation, the chemical and magnetic interactions between cations and spins are modelled using pair interaction energies out to third nearest neighbours. The distribution of the cations and orientations of their magnetic moments (either left or right in this case) are varied until thermodynamic equilibrium is achieved (that is, the free energy is minimized). Instantaneous snapshots of the equilibrium structure can then be examined. Cations are coloured according to chemistry and orientation of magnetic moment (red, Fe^{3+} , moment $5 \mu_B$ pointing right; orange, Fe^{3+} , moment $5 \mu_B$ pointing left; yellow, Fe^{2+} , moment $4 \mu_B$ pointing right; green, Fe^{2+} , moment $4 \mu_B$ pointing left; blue, Ti with no moment). Upper and lower parts of figure correspond to essentially pure antiferromagnetic haematite (alternating Fe^{3+} layers with left and right spins). In the centre is a lamella of essentially pure paramagnetic ilmenite (Ti layers alternating with layers containing equal numbers of Fe^{2+} -spin left and Fe^{2+} -spin right). Layers at interfaces between ilmenite and haematite contain a mixture of Fe^{2+} and Fe^{3+} . They do not correspond to the chemistry of either haematite or ilmenite, and are called ‘contact layers’. They were observed in all simulations performed within the low-temperature antiferromagnetic haematite + paramagnetic ilmenite stability field.

ferrimagnetic ferri-ilmenites have low Curie temperatures, and are paramagnetic above room temperature.

According to the above description, exsolution intergrowths of titanohaematite and ferri-ilmenite, such as those depicted in Fig. 1, contain an antiferromagnetic phase (with small canted moment) and a paramagnetic phase, neither of which has the correct combination of properties to explain the strong remanence observed in these materials. To determine the source of magnetization in these fine-scale intergrowths, an atomistic simulation of cation ordering, magnetic ordering, and exsolution in the ilmenite–haematite solid solution was performed, using the Monte Carlo method^{14,15}. Figure 2 shows a small section of a Monte Carlo simulation cell performed at 500 K and a bulk composition of 25 mol% ilmenite (Ilm₂₅). The upper and lower parts of Fig. 2 correspond to essentially pure antiferromagnetic haematite (layers of Fe³⁺ with alternating left and right spins). In the centre is a lamella of essentially pure paramagnetic ilmenite (layers of Ti alternating with layers containing equal numbers of Fe²⁺-spin left and Fe²⁺-spin right). The cation layers at the interface between ilmenite and haematite contain a mixture of Fe²⁺ and Fe³⁺. These layers do not correspond to the chemistry of either haematite or ilmenite, and are referred to as ‘contact’ layers. They improve, though not perfectly, local Pauling bond-strength satisfaction of oxygen in layers parallel to (001).

The presence of contact layers creates an imbalance between the equal and opposite α and β sublattice magnetizations of the antiferromagnetic haematite host. This imbalance generates a net ferrimagnetic moment, which necessarily adopts both the coercivity and thermal stability of the haematite host. This concept is best illustrated using simplified layer models (Fig. 3). Figure 3a and b

illustrates an antiferromagnetic host of composition Ilm₁₄ containing two paramagnetic ilmenite lamellae of composition Ilm₈₃. These compositions were chosen to represent possible low-temperature equilibrium states according to the calculated phase diagram¹⁶. The haematite host is antiferromagnetic, containing an equal number of left and right magnetic moments. Insertion of the lower ilmenite lamella removes an odd number of haematite layers, creating a net moment equivalent to the sublattice magnetization of a single haematite layer (M_{haem}), pointing to the right. The sublattice magnetizations of the two contact layers (M_{con}) are antiferromagnetically coupled to the haematite host, creating a magnetic moment of $2M_{\text{con}}$, pointing to the left. The end result is a net ferrimagnetic moment for the lamella of $2M_{\text{con}} - M_{\text{haem}}$, pointing to the left (large arrow).

The orientation of the moment associated with each lamella depends on its vertical position when nucleated. In Fig. 3a, the upper lamella is magnetically in-phase with the lower one, and the moments are additive. For the compositions shown, a net ferrimagnetic moment of $0.28 \mu_B$ per formula unit (p.f.u.) (where μ_B is the Bohr magneton) is achieved. In Fig. 3b, the upper lamella is magnetically out-of-phase with the lower one, and the moments cancel.

Figure 3c and d illustrates a paramagnetic host of composition Ilm_{89.7} containing two haematite precipitates of composition Ilm₆. Lamellae in Fig. 3c and d occupy identical positions, dictated by the surrounding ilmenite (that is, each haematite lamella must be initiated on a former Ti layer of the ilmenite host). In Fig. 3c, the moment of the upper lamella is parallel to the lower one, whereas in Fig. 3d it is antiparallel. Thus, the orientation of the moments

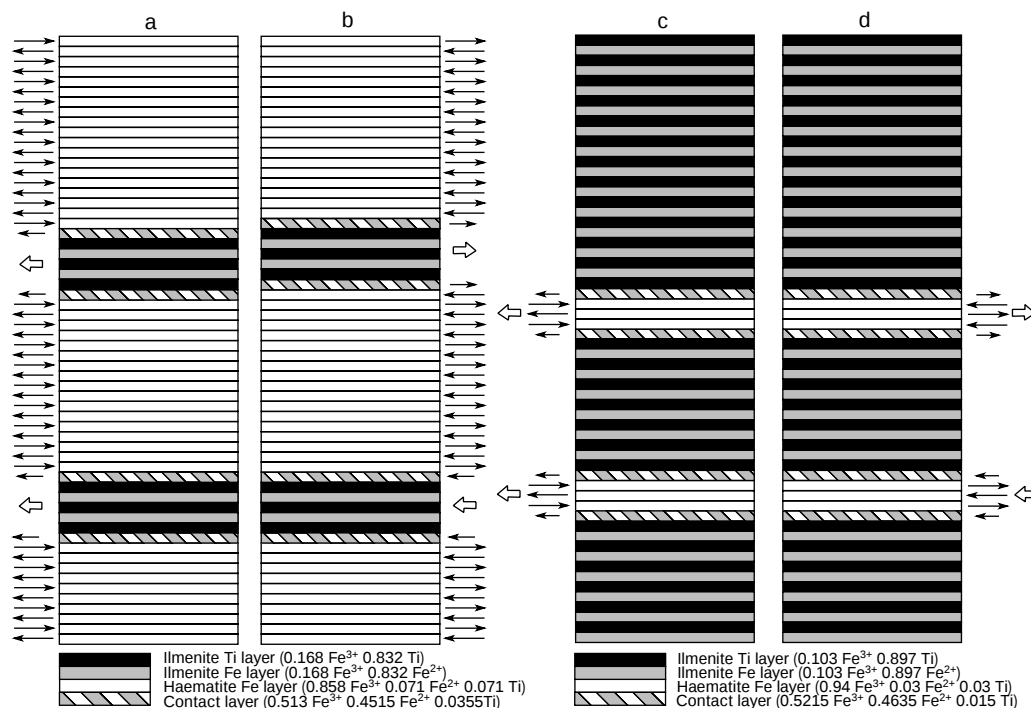


Figure 3 Models of multilayer lamellar magnetism. Panels **a** and **b** show haematite hosts, **c** and **d** show ilmenite hosts. Dark arrows indicate the direction and relative magnitude of magnetic moments in each cation layer. Open arrows indicate the direction of net moment for each lamella ($4.168 \mu_B$ in **a** and **b**; $4.103 \mu_B$ in **c** and **d**). **a, b**, 60 layers in a 20/80 ratio of paramagnetic ilmenite (Ilm_{83.2}) and antiferromagnetic haematite (Ilm_{14.2}). Each contains two ilmenite lamellae, one unit cell thick. In **a**, the moments of the two lamellae are in-phase with each other and the net moment is $(2 \times 4.168)/30 = 0.278 \mu_B$ p.f.u. In **b**, the moments of the two lamellae are out-of-phase with each other, and the net moment

is zero. Physical positioning of an ilmenite lamella determines the direction, left or right, of its net moment within an antiferromagnetic haematite host. **c, d**, 100 layers in a 92/8 ratio of ilmenite (Ilm_{89.7}) and haematite (Ilm₆). Bottoms of models have been cut to save space. Each contains two haematite lamellae, two-thirds of a unit cell thick. For **c**, the net moment of the upper lamella points to the left, whereas in **d** it points to the right. For **c**, the net moment is $(2 \times 4.103 \mu_B)/50 = 0.164 \mu_B$ p.f.u. For **d**, the net moment is zero.

associated with haematite lamellae in an ilmenite host are not related to their physical placement, as is the case for ilmenite lamellae in a haematite host (Fig. 3a and b). For the compositions shown, the net moment in Fig. 3c is $0.164 \mu_B$ p.f.u. The differences between haematite and ilmenite hosts in Fig. 3 reflect variations in coercivity. Demagnetization in ilmenite requires only demagnetization of haematite lamellae; in haematite, the entire host must be demagnetized.

Natural haematite and ilmenite hosts should contain combinations of in-phase and out-of-phase lamellae. No crystallographic feature favours in-phase over out-of-phase nucleation, but there is a natural condition favouring in-phase lamellae. When (001) planes of haematite or ilmenite are parallel to the magnetic field at nucleation, the field will strongly favour nucleation of in-phase lamellae, and strong remanence in (001). In a haematite host, this implies that the field actually determines the physical placement of the nucleating ilmenite lamellae. In an ilmenite host, however, the field only determines the magnetic orientation of haematite lamellae; physical position is dictated by chemistry. When (001) planes are normal to the field, or in the absence of a field, a random distribution of in- and out-of-phase lamellae results, producing zero net magnetization.

The intensity of lamellar magnetism is strongest when (1) the proportion of exsolved lamellar material is large, (2) lamellae are of minimal size and maximum abundance, maximizing the surface area of contact layers, and (3) host (001) planes are parallel to the magnetizing field, promoting in-phase magnetization. In favourable natural circumstances, a saturated magnetic moment of $0.3 \mu_B$ p.f.u. might be achieved with magnetic moments of all lamellae in-phase and aligned, corresponding to 55 kA m^{-1} , which may be compared to 480 kA m^{-1} for pure magnetite and 2.5 kA m^{-1} (ref. 15) for pure spin-canted antiferromagnetic haematite. Thus, a magnetization 22 times larger than pure haematite is achievable, while retaining the high coercivity and thermal stability of single-domain haematite. A rock containing 1% of similar material, but with only 6% of lamellae magnetically in phase, would have a magnetization of 33 A m^{-1} .

Kletetschka *et al.*¹⁷ have an alternative explanation for the strong and stable remanence of titanohaematite in haemo-ilmenite. They claim early formation, during cooling, of coarse antiferromagnetic haematite lamellae, which are multidomain and hence, according to these authors, capable of acquiring strong thermal remanent magnetization, which is merely 'hardened' by later exsolution. This would have a spin-canted magnetization perpendicular to the lamellar magnetic moment proposed here. However, in our interpretation of the phase diagrams^{15,16}, the only permissible coarse titanohaematite lamellae would be paramagnetic (magnetically disordered) during their formation, only becoming magnetized on cooling below 390°C (ref. 15) as they break down to a fine intergrowth of ordered antiferromagnetic haematite and cation-ordered ilmenite, qualitatively like the pattern illustrated in Fig. 3a. This low exsolution temperature shows that the magnetization is a chemical remanence, not a thermal remanent magnetization, because formation of the antiferromagnetic haematite takes place $100\text{--}200^\circ\text{C}$ below its Néel temperature.

Lamellar magnetism, with its 'hard' magnetic memory, is responsible for many remanence-dominated anomalies over ancient Earth rocks, and is a possible source for some deep-crustal magnetic anomalies on Earth or on planets such as Mars that once had a magnetic field. We believe that lamellar magnetism could also provide the basis for development of a highly stable magnetic storage unit. □

Received 17 May; accepted 26 June 2002; doi:10.1038/nature00942.

1. Ishikawa, Y. & Akimoto, S. Magnetic property and crystal chemistry of ilmenite (MeTiO_3) and hematite ($\alpha\text{Fe}_2\text{O}_3$) system II. Magnetic property. *J. Phys. Soc. Jpn* **13**, 1298–1310 (1958).
2. Balsley, J. R. & Buddington, A. F. Iron titanium oxide minerals, rocks, and aeromagnetic anomalies of

the Adirondack area, New York. *Econ. Geol.* **53**, 777–805 (1958).

3. Hargraves, R. B. Magnetic anisotropy and remanent magnetism in hemo-ilmenite ore deposits at Allard Lake, Quebec. *J. Geophys. Res.* **64**, 1565–1578 (1959).
4. Carmichael, C. M. The magnetic properties of ilmenite-hematite crystals. *Proc. R. Soc. Lond. A* **263**, 508–530 (1961).
5. Merrill, R. T. & Grommé, C. S. Nonreproducible self-reversal of magnetization in diorite. *J. Geophys. Res.* **74**, 2014–2024 (1969).
6. McEnroe, S. A. & Brown, L. L. A closer look at remanence-dominated aeromagnetic anomalies: Rock-magnetic properties and magnetic mineralogy of the Russell Belt microcline-sillimanite gneiss, northwest Adirondack Mountains, New York. *J. Geophys. Res.* **105**, 16437–16456 (2000).
7. McEnroe, S. A., Harrison, R. J., Robinson, P., Golla, U. & Jercinovic, M. J. Effect of fine-scale microstructures in titanohaematite on the acquisition and stability of natural remanent magnetization in granulite-facies metamorphic rocks southwest Sweden: Implications for crustal magnetism. *J. Geophys. Res.* **106**, 30523–30546 (2001).
8. McEnroe, S. A., Robinson, P. & Panish, P. T. Detailed chemical and petrographic characterization of ilmenite- and magnetic-rich cumulates of the Sokndal region, Rogaland, Norway. *Norweg. Geol. Surv. Bull.* **436**, 49–56 (2000).
9. McEnroe, S. A., Robinson, P. & Panish, P. T. Aeromagnetic anomalies, magnetic petrology and characterization of ilmenite- and magnetic-rich cumulates of the Sokndal region, Rogaland, Norway. *Am. Mineral.* **86**, 1447–1468 (2001).
10. Robinson, P., Panish, P. T. & McEnroe, S. A. Minor element chemistry of hemo-ilmenite- and magnetite-rich cumulates from the Sokndal region, south Rogaland, Norway. *Am. Mineral.* **86**, 1469–1476 (2001).
11. Banerjee, S. K. Magnetic properties of Fe-Ti oxides. *Rev. Mineral.* **25**, 107–128 (1991).
12. Dunlop, D. J. & Özdemir, O. *Rock Magnetism: Fundamentals and Frontiers* (Cambridge Univ. Press, New York, 1997).
13. McEnroe, S. A., Robinson, P. & Panish, P. T. Correlation between oxide mineralogy, rock-magnetic properties and aeromagnetic anomalies: Case study from the Sokndal Region, South Norway. *Eos* **79**, S362 (1998).
14. Harrison, R. J. Magnetic transitions in minerals. *Rev. Mineral.* **39**, 175–202 (2000).
15. Harrison, R. J. & Becker, U. in *Solid Solutions in Silicate and Oxide Systems* (ed. Geiger, C. A.) Vol. 3, 349–383 (Notes in Mineralogy, European Mineralogical Union, Eötvös Univ. Press, Budapest, 2001).
16. Burton, B. P. The interplay of chemical and magnetic ordering. *Rev. Mineral.* **25**, 303–321 (1991).
17. Kletetschka, G., Wasilewski, P. J. & Taylor, P. T. The role of hematite-ilmenite solid solution in the production of magnetic anomalies in ground and satellite-based data. *Tectonophysics* **347**, 166–177 (2002).
18. Merrill, R. T. A possible source for the coercivity of ilmenite-hematite minerals. *J. Geomag. Electr.* **20**, 181–185 (1968).

Acknowledgements

TEM imaging and TEM analytical studies at Institut für Mineralogie, Universität Münster, were done in cooperation with A. Putnis and U. Golla, and atomic simulations were carried out in cooperation with U. Becker. Electron microprobe studies at Univ. Massachusetts were completed with assistance from P.T. Panish and M.J. Jercinovic. S.A.M. and P.R. were supported by the strategic research fund of the Geological Survey of Norway and Bayerisches Geo-Institut. The work of S.A.M. on Swedish granulites was initiated during a Fellowship from the American-Scandinavian Foundation and from the Institute of Rock Magnetism, and her work in the Adirondacks was supported by the NSF. R.J.H. carried out his studies while a research fellow at the Institut für Mineralogie, Universität Münster.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.R. (e-mail: peter.robinson@ngu.no).

A 20-km-diameter multi-ringed impact structure in the North Sea

Simon A. Stewart* & Philip J. Allen†

* BP plc, Burnside Road, Dyce, Aberdeen AB21 7PB, UK

† Production Geoscience Ltd, Banchory AB31 5YR, UK

Most craters found on Earth are highly eroded, poorly preserved and only exposed on land^{1,2}. Here we describe a multi-ringed impact structure discovered in the North Sea from the analysis of three-dimensional seismic reflection data. The structure is 20 km in diameter, and has at least ten distinctive concentric rings located between 2 and 10 km from the crater centre. The

structure affects Cretaceous chalk and Jurassic shales, and is well preserved below several hundred metres of post-impact Tertiary strata, which constrains its age to be 60–65 Myr old. The formation of concentric ringed impact structures at this relatively small scale had not previously been thought possible, especially on the terrestrial planets^{1,3,4}. We have mapped the ring structures at a resolution of tens of metres both laterally and in depth, and show that the rings are fault-bounded graben structures, similar to fault arrays formed in low-strain-rate detachment tectonic settings⁵. Strata deeper than 500 m palaeodepth appear unfaulted, and we infer that the concentric ring structures may have accommodated post-impact extension towards the excavated crater, through detachment on weak layers within the chalk^{6,7}.

Three-dimensional (3D) seismic reflection data acquired for routine hydrocarbon exploration in the southern North Sea show a previously unrecognized concentric multi-ringed structure, 130 km from the English coast (1° 51' E, 54° 14' N). The results of preliminary mapping and interpretation are reported here. We term the structure 'Silverpit crater' after a nearby sea-floor channel⁸. The structure was gently folded by late Tertiary regional tectonics, and now lies at a depth range spanning 300 m to 1,500 m below the sea bed. Present-day water depth is 40 m. The seismic data reveal a number of distinctive aspects of the structure, such as a conical central peak buried inside a 3-km-diameter crater that is in turn bounded by a set of several inward-facing fault scarps. But the most striking aspect of this structure is the plan-view pattern of concentric rings forming the outer parts of the structure, mapped at the top of the Cretaceous (Fig. 1a).

Unequivocal recognition of terrestrial impact structures is usually dependent on outcrop-scale criteria such as shock metamorphism and shatter cones⁹. Data from drilled wells are necessary to meet these criteria in buried crater structures. Two hydrocarbon exploration wells passed through the Silverpit structure, 43/25-1 (drilled in 1984) and 43/24-3 (1993), at radial distances of 2.75 km and 7.38 km respectively (Fig. 1a). These wells were drilled to probe pre-Permian strata. Data recovered from the Cretaceous and Tertiary sections were restricted to sparsely sampled drill cuttings that contain no irregularities, according to drilling logs. There is no anomaly on regional gravity data at the location of the structure⁸. A

prominent regional magnetic lineament passes 30 km to the north of the structure⁸, but again there is no anomaly coincident with the structure itself, arguing against a volcanic origin. The stratigraphy of the southern North Sea includes Permian salt that can be several hundred metres thick⁸. The seismic data demonstrate, however, that apart from some signal degradation due to the overlying structural complexities, the Triassic and top Permian seismic reflectors are continuous underneath the Silverpit structure, ruling out salt diapirism as a cause (Fig. 2). But the Silverpit structure exhibits geometrical characteristics of an impact structure, albeit a very unusual one.

On terrestrial, silicate planets, there appears to be a regular progression of crater morphology from small simple craters, through central peak and peak ring craters, to large multi-ring basins³. These multi-ring basins have a small number (2–4) of well-defined rings formed by asymmetric scarps that down-throw towards the crater, and the spacing between rings is a substantial fraction of their diameter. Examples are known from the Moon¹⁰ (Oriente is the type example), Venus and Earth, and all are hundreds of kilometres in diameter. On icy moons, there appears to be an analogous progression in crater morphology with size. The largest craters are often also multi-ringed, though their detailed morphology is very different to Oriente-type multi-ringed basins. The type example is Valhalla on Callisto¹¹. Valhalla-type impact craters have many, closely-spaced rings (Fig. 1b). Individual rings appear as a mixture of ridges, grabens, and inwardly and outwardly facing asymmetric scarps. The radial extent of the rings is typically large in comparison with the apparent size of the initial impact^{4,11}. The geometry of the Silverpit structure is more closely analogous to the Valhalla structure than to the Oriente structure typical of terrestrial planets (Fig. 1). But as well as the surprising discovery of this type of structure on Earth, the Silverpit structure is also unusual in that it is significantly smaller than any of the known Valhalla-type structures^{4,11}.

Three-dimensional seismic coverage of the Silverpit structure allows the internal structure of a concentric multi-ringed impact structure to be mapped at high resolution—for the first time, to the best of our knowledge. The Silverpit structure can be divided into three distinctive structural zones according to radial distance (Fig. 2). Zone 1 is a bowl-shaped crater approximately 2,960 m diameter

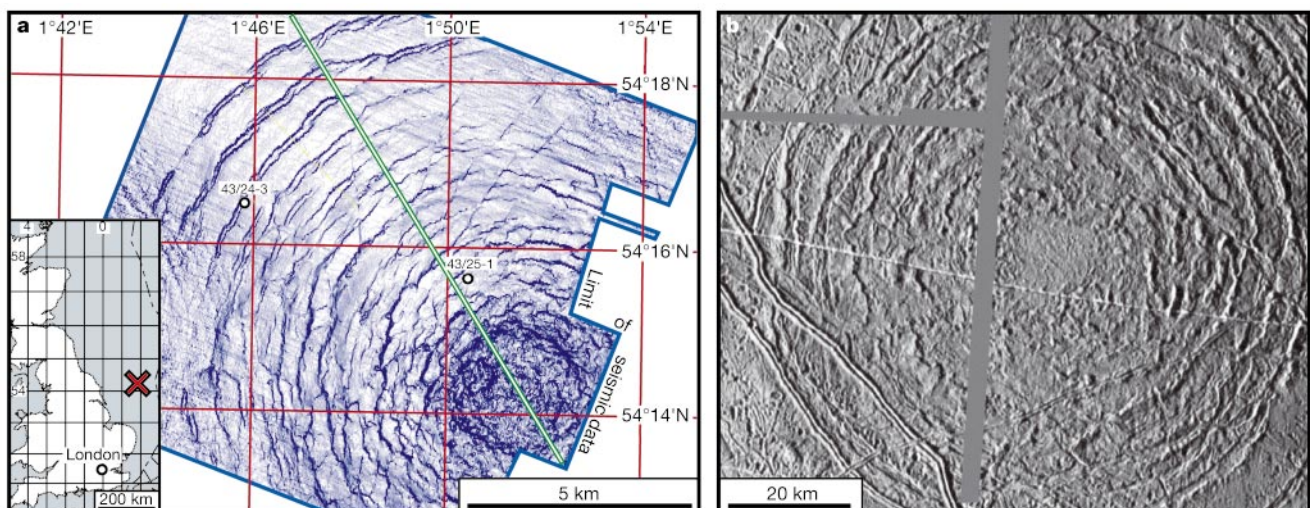


Figure 1 Silverpit crater compared with a similar concentric ringed structure on Europa. **a**, Silverpit crater mapped at top Cretaceous chalk. Bluescale shading is geological dip—darker is steeper, darkest linear trends are fault scarps. Numbered locations are hydrocarbon exploration wells. Green line is the location of the seismic line shown in Fig. 2.

Inset, location of study area. **b**, Tyre impact structure, Europa. Note the difference in scale. Target stratigraphy is believed to be an ice crust approximately 10 km thick, underlain by an ocean of water or brine⁴.

from rim to rim, excavated into the Cretaceous chalk, and is characterized by obliteration of the top Cretaceous seismic reflector. The best estimate of top Cretaceous within the crater gives a depth of 100–200 m below regional elevation (Fig. 2). The base Cretaceous chalk reflector is more clearly imaged, and defines a conical peak that was entirely buried below the crater floor (Figs 2 and 3). Measured at base Cretaceous, it is approximately 750 m in diameter and 250 m high. There are some minor radial faults associated with this peak (Fig. 3), but the Jurassic isopach (interval thickness) generally varies smoothly into the core of the peak, suggesting that peak growth was achieved by flow or a volume expansion mechanism such as irreversible fracturing¹². Triassic reflectors imaged 1 km below the crater appear unfaulted, though there is seismic signal degradation and velocity push down.

Zone 2 lies between approximately 1.5 km and 4 km from the crater centre, and is characterized by a set of concentric extensional fault scarps that face the crater and define a set of tilted fault blocks (Fig. 2). The faults are evenly spaced with separations of 300–500 m in the facing direction, and maximum displacements of approximately 50 m. The faults are well developed at the top of the Cretaceous, but do not offset the base Cretaceous. Some definitions of crater size measure diameter between the scarps that bound the

outermost extensional fault terrace³. This definition would give a diameter of approximately 8 km for the Silverpit crater. We restrict our estimate of crater diameter to the central crater of zone 1, owing to the distinctive crater morphology and obliterated seismic facies of that zone compared with the seismic character in zone 2, which exhibits the intact (but faulted) regional character of top Cretaceous chalk. The fault heaves record 8% line length extension of this zone.

Zone 3 lies between approximately 4 km and 10 km from the crater centre, and is characterized by rings of concentric fault-bound graben (Fig. 4). In plan view, the graben show minor offsets and relay ramps, giving the fault system a similar geometrical appearance to systems formed at relatively low strain rates⁵. In cross-section, the graben-bounding faults, like the faults of zone 2, terminate in the lower half of the Cretaceous sequence (Fig. 4). The faults are steeper than those in zone 2, so represent less extension towards the central crater. The graben present a geometrical space problem in terms of how volume reduction is accommodated within and below the lower parts of the downfaulted keystone blocks.

The seismic stratigraphic architecture of the Silverpit structure ties to the loosely defined chronostratigraphy of well 43/25-1⁸ (Fig. 2). The tops of the tilted fault blocks in zone 2 define a surface that

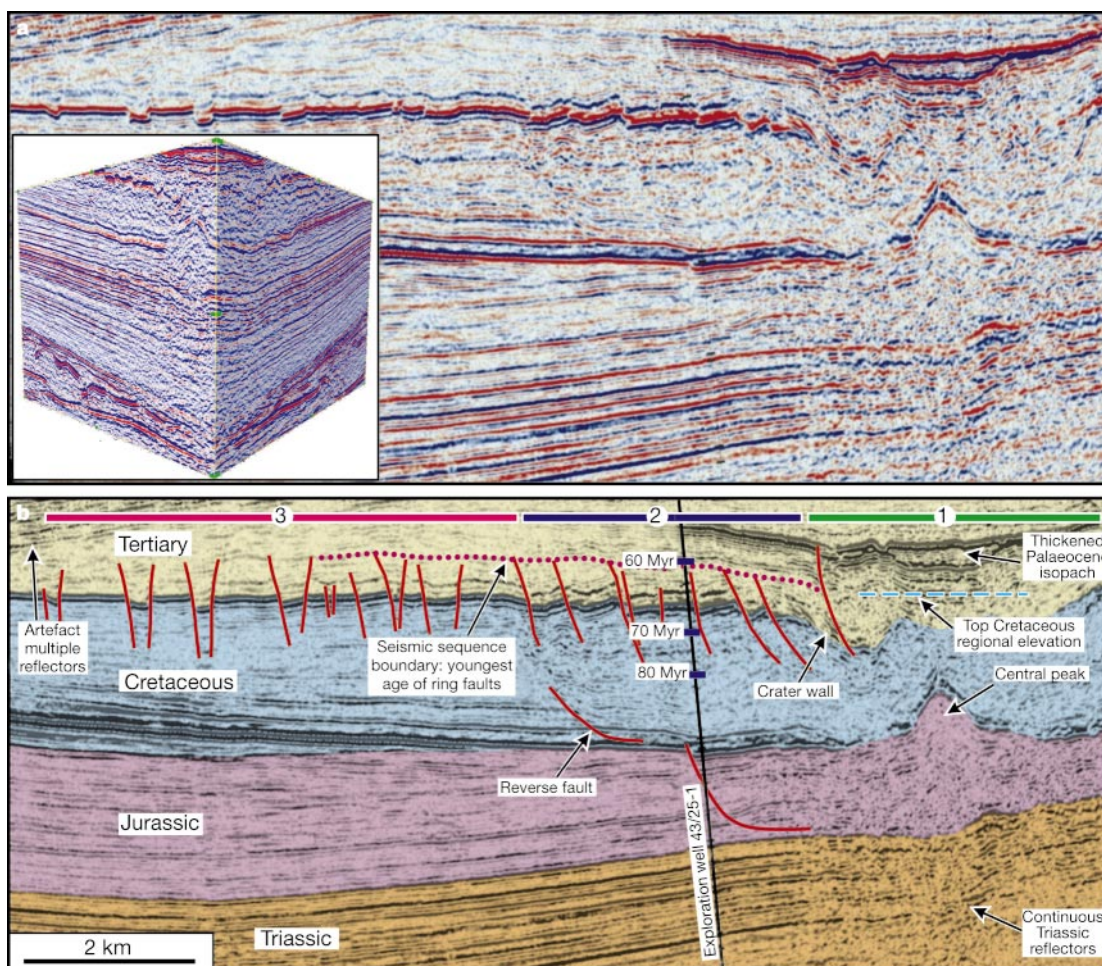


Figure 2 Seismic examples from the Silverpit structure, showing the structural form and geometry of the older and younger strata. **a**, Uninterpreted seismic line, location shown on Fig. 1. Inset cube shows the relatively unstructured underlying Triassic and top Permian seismic reflectors. **b**, Annotated version of seismic line. The vertical scale of seismic and annotated sections is exaggerated approximately 2×. Zones 1, 2 and 3 respectively correspond to the central crater, and zones of concentric rings characterized by

asymmetric tilted fault blocks and symmetrical graben. Key features of the central crater include the obliterated nature of the top Cretaceous seismic reflector, showing that the crater post-dates deposition of the youngest chalk, and the prominent peak feature within the crater. The slight depression of the Triassic reflectors (velocity push down) below the crater indicates that there is a local low seismic velocity anomaly in the vicinity of the zone 1 crater.

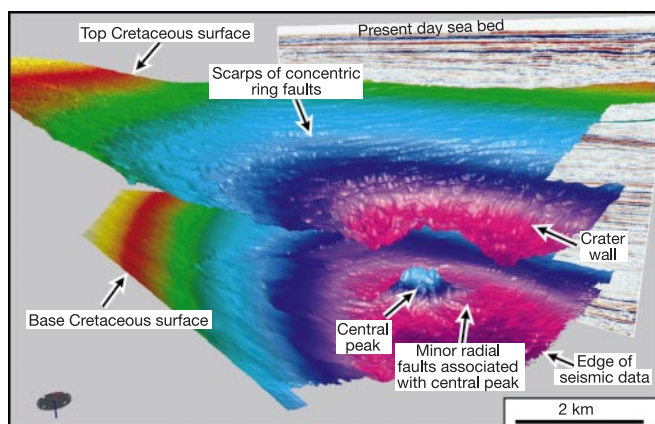


Figure 3 Perspective view, showing the form of the mapped structure in three dimensions. View is from the southeast. The buried central peak developed at the base of the Cretaceous lies below the crater mapped at the top Cretaceous. Surface colour corresponds to present-day depth below sea bed. A seismic line has been placed at the back of the displayed surfaces to illustrate the source data characteristics. The top Cretaceous surface has been cut away slightly to enable the underlying central peak to be clearly seen.

ties to well 43/25-1 at intra-Palaeocene level, with an age of approximately 60 Myr (Fig. 2), though it is difficult to determine whether the faults represent instantaneous displacements or whether there was prolonged movement after creation of the crater. Nonetheless, the structure cannot be younger than the zone 2 fault blocks, but it is younger than the chalk, the top of which is obliterated in zone 1. As the youngest chalk is Maastrichtian in this area⁸, the age of the structure is constrained to the range 60–65 Myr. The Tertiary sediments above zone 1 show late, conical extensional faults and isopach thickening, both indicating post-impact differential compaction that has also been noted in association with other marine impact structures¹³. Sediment facies indicate marine water depths in the Silverpit area of 50–300 m throughout the timespan of 60–65 Myr ago⁸.

Mechanical models of multi-ringed craters suggest that a prerequisite for the formation of concentric rings is a decrease in target strength with depth^{4,6,7}. The strength profile of the Cretaceous

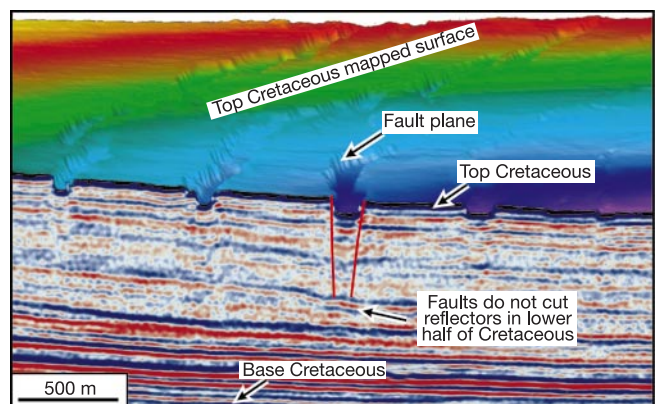


Figure 4 Detailed view of the seismic structure and interpretation, illustrating the three-dimensional form of the concentric ring structures. Data taken from zone 3 of Silverpit structure. Top Cretaceous surface is coloured according to present-day depth. Depth of view from front to back of displayed surface is approximately 4 km. The graben form of the concentric rings is clear; the seismic panel shows that the graben-bounding faults penetrate to approximately halfway through the Cretaceous. Vertical exaggeration of seismic structure and displayed surface is approximately 2x the horizontal scale.

sediments in the Silverpit area at the time of impact is poorly constrained, but regional studies indicate that the lower parts of the Cretaceous chalk are interbedded with clays¹⁴. We suggest that differential compaction of clay interbeds relative to chalk ooze, combined with the high primary porosity of the ooze (up to 70%), led to slightly overpressured chalk layers that acted as detachments in the lower parts of the Cretaceous sequence during the formation of the Silverpit structure. Modelling indicates that ring fractures reach a similar depth to the maximum penetration of the impactor¹². In combination with scaling relationships for transient crater dimensions^{1,2}, this suggests that ring fractures around the Silverpit crater propagated to 300–600 m below the sea floor (tying broadly with the observed lower fault tips in zone 3). Penetration of fractures to the level of the overpressured chalk layers at the time of impact would have provided dewatering conduits, solving the volume accommodation problem that otherwise exists below the downfaulted graben of zone 3 (Fig. 4). Detachments would have accommodated post-impact gravitationally driven collapse towards the crater in zones 2 and 3 over a geological time span that in turn could account for the similarity between the final concentric ring geometries and other relatively low-strain-rate fault systems. □

Methods

Seismic data acquisition and processing

The 3D marine seismic data were acquired in 1992 by Western Geophysical on behalf of Arco International Oil and Gas Company, and are now owned by BP plc. A dual-streamer, dual-source (flip-flop) configuration was used. Each source was a 2,250 inch³ sleeve gun array. Each streamer consisted of 288 channels spaced every 12.5 m. Common depth point line spacing was 25 m with nominal 36-fold coverage. The data were reprocessed in 1996 by Compagnie Générale de Géophysique, using a standard processing sequence that included post-stack 3D time migration. The crossline trace spacing was interpolated, so that the final sampling density and data resolution was 12.5 m by 12.5 m. Vertical sampling was 4 ms (approximately 6 m).

Received 20 March; accepted 17 June 2002; doi:10.1038/nature00914.

- Grieve, R. A. F. & Pesonen, L. J. The terrestrial impact cratering record. *Tectonophysics* **216**, 1–30 (1992).
- Grieve, R. A. F. Extraterrestrial impacts on earth: the evidence and the consequences. *Geol. Soc. Lond. Spec. Publ.* **140**, 105–131 (1998).
- Melosh, H. J. *Impact Cratering: A Geological Process* (Oxford Univ. Press, New York, 1989).
- Moore, J. M. *et al.* Large impact features on Europa: Results of the Galileo nominal mission. *Icarus* **135**, 127–145 (1998).
- Schultz-Elia, D. D. & Walsh, P. Modeling of grabens extending above evaporites in Canyonlands National Park, Utah. *J. Struct. Geol.* **24**, 247–275 (2002).
- Melosh, H. J. & McKinnon, W. B. The mechanics of ringed basin formation. *Geophys. Res. Lett.* **5**, 985–988 (1978).
- Allemand, P. & Thomas, P. Small-scale models of multiring basins. *J. Geophys. Res.* **E 104**, 16501–16514 (1999).
- Cameron, T. D. J. *et al.* *United Kingdom Offshore Regional Report: The Geology of the Southern North Sea* (HMSO, London, 1992).
- French, B. M. 25 years of the impact-volcanic controversy: Is there anything new under the sun or inside the Earth? *Eos* **71**, 411–414 (1990).
- Spudis, P. D., Reisse, R. A. & Gillis, J. J. Ancient multiring basins on the moon revealed by Clementine laser altimetry. *Science* **266**, 1848–1851 (1994).
- Greeley, R., Kiemaszewski, J. E. & Wagner, R. Galileo views of the geology of Callisto. *Planet. Space Sci.* **48**, 829–853 (2000).
- O'Keefe, J. D., Stewart, S. T., Lainhart, M. E. & Ahrens, T. J. Damage and rock-volatile mixture effects on impact crater formation. *Int. J. Impact Eng.* **26**, 543–553 (2001).
- Tsikalas, F., Gudlaugsson, S. T. & Faleide, J. I. Collapse, infilling, and postimpact deformation at the Mjølir impact structure, Barents Sea. *Geol. Soc. Am. Bull.* **110**, 537–552 (1998).
- Wray, D. S. & Wood, C. J. Distinction between detrital and volcanogenic clay-rich beds in Turonian–Coniacian chalks of eastern England. *Proc. Yorks. Geol. Soc.* **52**, 95–105 (1998).

Acknowledgements

Published with permission of BP plc and PGL Ltd. The views expressed here are solely those of the authors and not necessarily those of BP plc or PGL Ltd.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.A.S. (e-mail: stewarsa1@bp.com).

The evolution of inaccurate mimics

Rufus A. Johnstone

Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Instances of strikingly accurate Batesian mimicry (in which a palatable prey organism closely resembles an aversive model) are often cited to illustrate the power of natural selection¹. Less attention has been paid to those mimics, such as many hoverfly (Syrphidae) mimics of wasps or bees, that resemble their models only poorly^{2–4}. Attempts to provide an adaptive explanation for imperfect mimicry have suggested that what seems a crude resemblance to human observers may appear a close match to predators², or that inaccurate mimics may bear a general resemblance to several different model species^{3,4}. I show here, however, that truly inaccurate mimicry of a single model organism may be favoured over perfect resemblance, by kin selection. Signal detection theory predicts that predators will modify their level of discrimination adaptively in response to the relative frequencies and similarity of models and mimics^{5–7}. If models are rare and/or weakly aversive, greater local similarity of mimics can thus lead to greater attack rates. Where individual mimics are related to others in their vicinity, kin selection will then oppose the evolution of accurate mimicry.

Existing explanations for imperfect mimicry are motivated by the common assumption that closer resemblance (from the predator's perspective) between a mimic and an aversive model cannot but reduce the former's risk of attack. Here, I show that accurate mimics may actually suffer greater rates of predation than individuals that resemble the same model less closely. My argument is based on the assumption that members of the mimetic population tend to be (to some degree) related to others in their vicinity. This kind of family grouping has been invoked in many discussions of the evolution of aposematism^{8–11}, but has received much less attention in the context of mimicry. As I demonstrate, however, it may lead to kin selection opposing the evolution of accurate resemblance.

Consider a population of Batesian mimics that resemble (to some degree) an unpalatable model species that is found in the same area. We wish to assess the impact of selection on the accuracy of the mimic population. To do so, we must determine the predation risk suffered by an individual mimic that resembles the model species more or less closely than is typical.

Suppose, then, that local predators learn to identify and avoid the distasteful models they encounter—that is, they learn the cues characteristic of the model species, forming a template with which they can compare the phenotypes of individual prey. Any particular model is likely to match a predator's template more closely than is an imperfect mimic; but there will be some degree of overlap between the probability distributions of perceived dissimilarity to the template for the two prey types. Consequently, a predator cannot simultaneously maximize its chances of avoiding models and of attacking mimics. To determine predation risk, we must determine the optimal trade-off for predators between the risk of inappropriately attacking a model and of inappropriately avoiding a mimic. This trade-off will depend upon the costs of the two types of error, on the relative frequencies of models and mimics and on the locally typical degree of dissimilarity between the two prey types^{5–7}.

Formally, I will assume that the perceived dissimilarity d_p between a predator's template and the appearance of a prey is normally distributed with standard deviation σ (>0) and with mean 0 for an individual model and $d\sigma$ for an individual mimic (where $d(\geq 0)$ represents the actual dissimilarity of the latter to the model

species), as illustrated in Fig. 1. Assuming that attacking a mimic yields a benefit b to the predator and attacking a model yields a cost c (avoiding either the mimic or the model yields a zero payoff), the predator then attempts to attack an individual prey item if its perceived dissimilarity from the template exceeds some critical threshold, the value of which depends on the relative frequencies of models ($1 - p$) and of mimics (p), and on the locally typical dissimilarity between the two, $\bar{d}$. This optimal threshold is equal to $t^*\sigma$, defined by

$$t^* = \frac{\bar{d}}{2} - \frac{K}{\bar{d}}, \text{ where } K = \ln \left[\frac{pb}{(1-p)c} \right] \quad (1)$$

The parameter K may be viewed as the 'incentive to attack'; a high value indicates that mimics are common and/or that the costs of attacking a model are small (relative to the benefits of attacking a mimic).

If predators adopt this optimal threshold (based on the typical

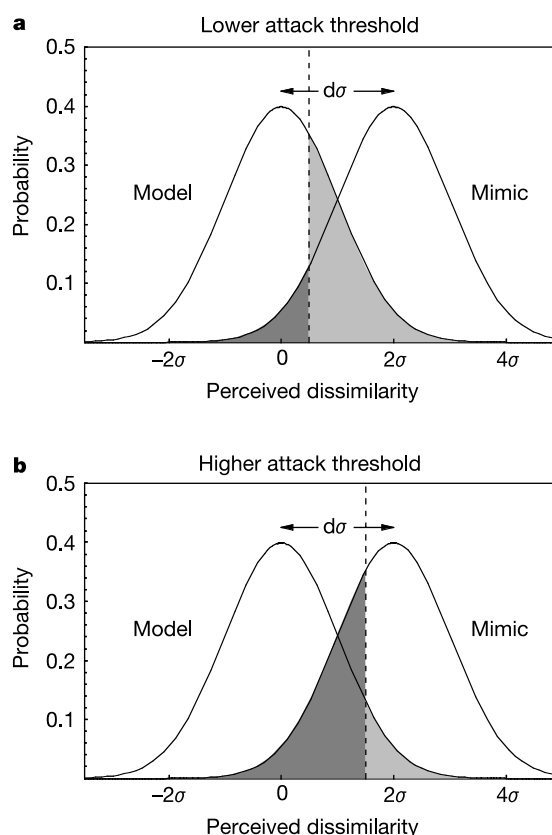


Figure 1 Predator discrimination. The graphs show the probability distributions of perceived dissimilarity to the predator's template for a model and for a mimic (when d , the mimic's actual degree of dissimilarity to the model, is equal to 2). The dotted lines represent the threshold levels of dissimilarity above which the predator will attack. The dark and light shaded areas give the probabilities of the two kinds of error, avoiding a mimic and attacking a model. **a**, The predator sets a low threshold; this reduces the chances of avoiding a mimic, but increases the chances of attacking a model. **b**, The predator sets a higher threshold, which has the opposite effect. Assuming that attacking a mimic yields a benefit b to a predator and attacking a model yields a cost c (while avoiding either yields a zero payoff), the expected payoff gained from an encounter in which the predator sets an attack threshold of $t\sigma$ is given by $pb(1 - Z(t - \bar{d})) - (1 - p)c(1 - Z(t))$, where p is the frequency of mimics, $1 - p$ the frequency of models, $\bar{d}$ is the locally typical degree of dissimilarity between models and mimics and Z is the cumulative normal distribution function with mean 0 and standard deviation 1. Maximizing this expression with respect to t yields the optimum attack threshold specified by (1) in the main text (the assumption that the predator maximizes its payoff per encounter is justified when handling times are similar for both models and mimics⁹).

characteristics of local prey), then the risk of attack suffered by a mimic of dissimilarity d , when the locally typical degree of dissimilarity between models and mimics is $\bar{d}$, is equal to

$$1 - Z\left(\frac{\bar{d} - K}{\bar{d}} - d\right) \quad (2)$$

where Z denotes the cumulative normal distribution function with mean 0 and standard deviation 1 (the above expression represents the probability of attack by an individual predator that encounters the mimic).

Now suppose that individuals tend to be related to others in their vicinity. Kin are likely to share the same phenotype, so accurate mimics will tend to be surrounded by neighbours that also resemble the model species closely. Conversely, inaccurate mimics will tend to be surrounded by neighbours that also resemble the model species poorly. To determine the impact of kin selection on the accuracy of the mimic population, we must take into account the influence of locally typical similarity on predation risk.

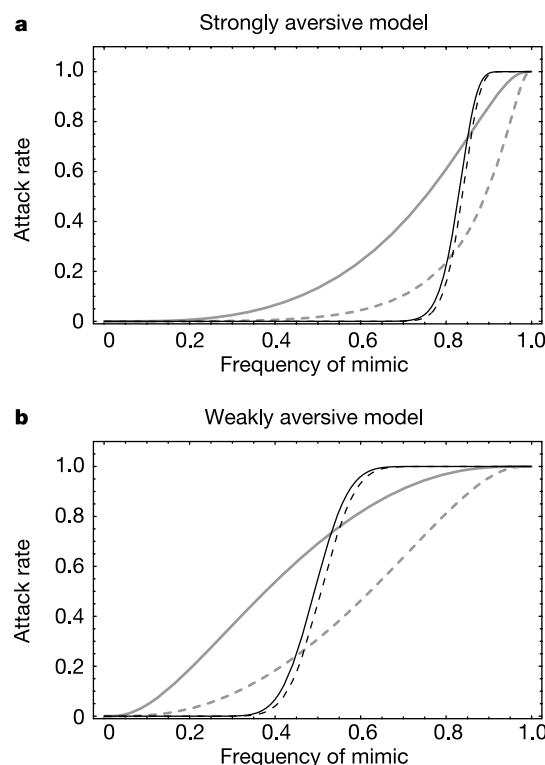


Figure 2 Impact of the frequency and accuracy of mimics on predator attack rates. The graphs show the probability of attack by predators for models (dotted lines) and for mimics (solid lines), as a function of the relative frequency of the latter, given that predators adopt the optimal attack threshold specified by equation (1) in the main text. The black lines in each graph are for a population of relatively accurate mimics ($\bar{d} = 0.25$); the grey lines are for a population of relatively inaccurate mimics ($\bar{d} = 1$). **a**, Results for the case of a more strongly aversive model ($c/b = 5$); **b**, results for a more weakly aversive model ($c/b = 1$). The figure indicates that an increase in the frequency of mimics typically leads to a decline in their protection from attack, and that this trend is initially more pronounced for inaccurate mimics. Accurate mimics may continue to enjoy a high level of deterrence even as they attain substantial frequencies (provided that the model is sufficiently aversive). Above a critical frequency, however, there is a rapid collapse in their level of protection (and in that enjoyed by the model), so that they suffer even higher levels of attack than do less accurate mimics^{6,7}. These are the circumstances under which kin selection may favour less precise resemblance between mimic and model. The less strongly aversive the model, the lower the critical frequency above which imperfect mimicry may be favoured.

Formally, employing the approach of Frank¹², individual fitness can be written as a function of individual and locally typical dissimilarity to the model, $w(d, \bar{d})$. The evolutionarily stable level of similarity d^* then satisfies

$$\frac{\partial w(d, \bar{d})}{\partial d} + r \frac{\partial w(d, \bar{d})}{\partial \bar{d}} = 0 \quad \text{for } d = \bar{d} = d^* \quad (3)$$

where $r (= \partial \bar{d} / \partial d)$ is a kin selection coefficient that specifies the regression of local average similarity on the similarity of the focal mimic (the appropriate neighbourhood over which one should calculate local relatedness will depend upon the spatial scale over which predators adjust their foraging behaviour). Equation (3) implies that, at equilibrium, the marginal fitness consequences of an increase in an individual's own similarity to the model must be balanced by the marginal fitness consequences of an increase in the locally typical similarity of neighbours to the model (weighted according to the strength of relationship between individual and locally typical similarity).

Equation (2) indicates that predation risk will always decrease with an individual mimic's similarity to the model (that is, risk will increase with d). In other words, individual selection always favours accurate mimicry. The effect of the locally typical degree of similarity, by contrast, depends on K and $\bar{d}$. When

$$K + (\bar{d}^2/2) < 0 \quad (4)$$

that is, when predators' incentive to attack is low ($K < 0$) and local mimics are not too dissimilar to the model species, individual predation risk decreases with locally typical similarity (that is, the marginal change in attack probability with respect to $\bar{d}$ is positive)¹³. Under these circumstances, kin selection reinforces individual selection, because more accurate mimics benefit by virtue of the similarity of their own close resemblance). But when condition (4) is not met, for example, when predators' incentive to attack is high (because mimics are common and/or models only weakly aversive), predation risk increases with locally typical similarity (that is, the marginal change in attack probability with respect to $\bar{d}$ is negative). Under these circumstances, kin selection opposes individual selection. The reason is that increasingly accurate mimicry among the local population leads to less cautious attack by predators (see Fig. 2). Consequently, a more accurate mimic may benefit because it resembles the model closely itself, but will suffer because its neighbouring relatives also do so.

What are the net consequences of individual selection and kin selection? Combining (2) and (3) (assuming that predators adopt the optimal attack threshold given the typical characteristics of local prey, and that prey fitness is a strictly decreasing function of predation risk) yields an evolutionarily stable level of dissimilarity d^* given by:

$$d^* = \begin{cases} 0, & \text{for } K \leq 0 \\ \sqrt{\frac{2Kr}{2-r}}, & \text{for } K > 0 \end{cases} \quad (5)$$

Equation (5) implies that when K is negative (that is, when models are common and/or strongly aversive) or when $r = 0$, mimics are expected to evolve ever closer resemblance to their model. Under these circumstances, kin selection either reinforces individual selection or opposes it too weakly to favour inaccurate mimicry. By contrast, when K and r are positive (that is, when models are rare and/or weakly aversive), kin selection opposes individual selection sufficiently strongly to yield an equilibrium level of dissimilarity between mimic and model that is greater than zero. Higher levels of local relatedness and greater incentives to attack both favour greater dissimilarity to the model species. The present analysis thus provides an adaptive hypothesis to account for the existence of inaccurate mimics.

When K and r are positive, and mimics do not evolve perfect resemblance to the model species, the resulting probabilities of predator deterrence by models and by mimics are given, respectively, by

$$Z\left(-(1-r)\sqrt{\frac{2K}{r(2-r)}}\right) \quad (6a)$$

and

$$Z\left(-\sqrt{\frac{2K}{r(2-r)}}\right) \quad (6b)$$

(I note that the probability of predator deterrence is equal to one minus the probability of attack). In order to quantify the accuracy of mimicry in a measurable manner, Fig. 3 shows the equilibrium probability of predator deterrence by mimics relative to the probability of deterrence by models, for a range of different parameter values.

The present hypothesis yields a number of predictions that might be tested either experimentally, or through comparative analysis of the distribution of inaccurate mimicry among prey species. First, it

predicts (as illustrated in Fig. 2) that a population of accurate mimics will enjoy greater protection than a population of inaccurate mimics when models are common and/or strongly aversive, but may suffer more frequent attack when models are sufficiently rare and/or only weakly aversive (because, under these circumstances, greater accuracy leads to less cautious attack by predators). Second, it predicts (as illustrated in Fig. 3) that inaccurate mimicry will be favoured in species that resemble a weakly aversive model, and that are abundant relative to that model. Furthermore, inaccuracy will be more strongly favoured in species with limited dispersal and family grouping, which feature higher levels of local relatedness.

As yet, there are insufficient data to evaluate the first prediction, because relatively few experimental studies have set out to investigate the consequences of imperfect mimicry^{5,14–17}. Existing evidence does suggest that mimics gain less protection when they are relatively common compared to models (particularly when the latter are only weakly aversive)^{14–16,18–21}. Moreover, there is also some suggestion that imperfect mimics suffer a more rapid initial decline in protection as they become more common, while perfect mimics may continue to enjoy a high level of deterrence even as they attain substantial frequencies^{14,16,18} (see Fig. 2). But no studies have compared the attack rates suffered by accurate and inaccurate mimics at frequencies high enough to bring about a significant loss of deterrence in the case of accurate mimicry. Turning to the comparative predictions, however, the ‘paradoxical’ observation that poor mimics tend to be more abundant than accurate mimics and than their models^{2,4} accords well with the present theory. □

Received 12 December 2001; accepted 24 April 2002; doi:10.1038/nature00845.

1. Wickler, W. *Mimicry in Plants and Animals* (Weidenfeld & Nicholson, London, 1968).
2. Dittrich, W., Gilbert, F., Green, P., McGregor, P. & Grewcock, D. Imperfect mimicry: a pigeon's perspective. *Proc. R. Soc. Lond. B* **251**, 195–200 (1993).
3. Howse, P. E. & Allen, J. A. Satyrism: the evolution of apparent imperfection. *Proc. R. Soc. Lond. B* **257**, 111–114 (1994).
4. Edmunds, M. Why are there good and poor mimics? *Biol. J. Linn. Soc.* **70**, 459–466 (2000).
5. Duncan, C. J. & Sheppard, P. M. Sensory discrimination and its role in the evolution of Batesian mimicry. *Behaviour* **24**, 269–282 (1965).
6. Getty, T. Discriminability and the sigmoid functional response: how optimal foragers could stabilise model-mimic complexes. *Am. Nat.* **125**, 239–256 (1985).
7. Greenwood, J. J. D. Crypsis, mimicry and switching by optimal foragers. *Am. Nat.* **128**, 294–300 (1986).
8. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
9. Gittleman, J. L. & Harvey, P. H. Why are distasteful prey not cryptic? *Nature* **286**, 149–150 (1980).
10. Harvey, P. H. & Paxton, R. J. The evolution of aposematic coloration. *Oikos* **37**, 391–393 (1981).
11. Guilford, T. The evolution of conspicuous coloration. *Am. Nat.* **131**, S7–S21 (1988).
12. Frank, S. A. *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, 1998).
13. Rodríguez-Gironés, M. A. & Lotem, A. How to detect a cuckoo-egg: a signal-detection theory model for recognition and learning. *Am. Nat.* **153**, 633–648 (1999).
14. Pilecki, C. & O'Donald, P. The effects of predation on artificial mimetic polymorphisms with perfect and imperfect mimics at varying frequencies. *Evolution* **25**, 365–370 (1971).
15. Hetz, M. & Slobodkinoff, C. N. Predation pressure on an imperfect Batesian mimicry complex in the presence of alternative prey. *Oecologia* **76**, 570–573 (1988).
16. Lindström, L., Alatalo, R. V. & Mappes, J. Imperfect Batesian mimicry—the effects of the frequency and the distastefulness of the model. *Proc. R. Soc. Lond. B* **264**, 149–153 (1997).
17. Mappes, J. & Alatalo, R. V. Batesian mimicry and signal accuracy. *Evolution* **51**, 2050–2053 (1997).
18. Brower, J. V. Z. Experimental studies of mimicry. IV. The reactions of starlings to different proportions of models and mimics. *Am. Nat.* **94**, 271–282 (1960).
19. Lea, R. G. & Turner, J. R. G. Experiments on mimicry: II. The effect of a Batesian mimic on its model. *Behaviour* **42**, 131–151 (1972).
20. Goodale, M. A. & Sneddon, I. The effect of distastefulness of the model on the predation of artificial Batesian mimics. *Anim. Behav.* **25**, 660–665 (1977).
21. Pfennig, D. W., Harcombe, W. R. & Pfennig, K. S. Frequency-dependent Batesian mimicry. *Nature* **410**, 323 (2001).

Acknowledgements

I thank R. Bshary, S. Dall, T. Getty, A. Lotem, S. Rands and A. Russell for comments and discussion.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to the author (e-mail: raj1003@hermes.cam.ac.uk).

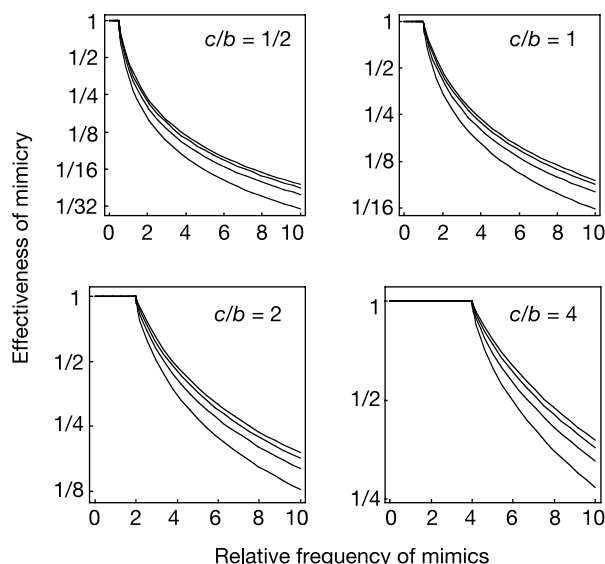


Figure 3 Equilibrium probability of deterrence by mimics relative to deterrence by models. Each graph shows relative deterrence by mimics, as a function of the ratio of mimics to models, for several different values of local relatedness; successively lower curves correspond in each case to successively higher values of $r = 1/16, 1/8, 1/4$ and $1/2$. The four graphs show results for different values of c/b : $1/2, 1, 2$ and 4 . When $r = 0$ (or $K \leq 0$), the equilibrium outcome of the model is perfect mimicry, so that mimics are precisely as effective (or ineffective) as models in deterring attack. When r and K are both positive, however, mimics are less effective at equilibrium than are models. Under these circumstances, a rare, perfectly mimetic type arising in the equilibrium population will enjoy less effective deterrence than typical (imperfectly mimetic) individuals. Taking into account local relatedness (which implies that the neighbours of a perfectly mimetic mutant will exhibit a lower average dissimilarity to the model than will the neighbours of a typical individual), and assuming that predators adopt the optimal threshold based on the average characteristics of local prey, the probability of deterrence by a mutant perfect mimic in the equilibrium population is, from equation (2) in the main text, equal to $Z[(1-r)d^*/2 - K/(1-r)d^*]$. To provide some illustrative results, the ratio of this value to the probability of deterrence by typical individuals (given by equation (6b) in the main text) is, for $c/b = 1$ and a 2:1 ratio of mimics to models, equal to 0.977, 0.951, 0.748 and 0.354 for $r = 1/32, 1/16, 1/8$ and $1/2$, respectively (all values quoted to three decimal places). These results indicate that even at low relatedness levels, a rare perfectly mimetic type may suffer a significant proportional reduction in deterrence compared to typical individuals.

Fractal geometry predicts varying body size scaling relationships for mammal and bird home ranges

John P. Haskell*, Mark E. Ritchie† & Han Olff‡

* College of Natural Resources, Utah State University, Logan, Utah 84322, USA

† Department of Biology, Syracuse University, Syracuse, New York 13244, USA

‡ Tropical Nature Conservation and Vertebrate Ecology Group, Wageningen University, Bornsesteeg 69, 6708 PD Wageningen, The Netherlands

Scaling laws that describe complex interactions between organisms and their environment as a function of body size offer exciting potential for synthesis in biology^{1–4}. Home range size, or the area used by individual organisms, is a critical ecological variable that integrates behaviour, physiology and population density and strongly depends on organism size^{5–7}. Here we present a new model of home range–body size scaling based on fractal resource distributions, in which resource encounter rates are a function of body size. The model predicts no universally constant scaling exponent for home range, but defines a possible range of values set by geometric limits to resource density and distribution. The model unifies apparently conflicting earlier results and explains differences in scaling exponents among herbivorous and carnivorous mammals and birds^{5–18}. We apply the model to predict that home range increases with habitat fragmentation, and that the home ranges of larger species should be much more sensitive to habitat fragmentation than those of smaller species.

Scaling relationships quantify the universal properties of complex physical and biological systems¹. These relationships can help identify patterns across different levels of biological organization, such as cells, individuals, populations and communities. For example, physiological characteristics of organisms, Y , often vary with body size, M , according to power functions of the general form:

$$Y = Y_0 M^b \quad (1)$$

where b is a scaling exponent, and Y_0 is a taxon- and character-specific normalization constant^{2–4}. Many important physiological rates seem to exhibit constant exponents b that are multiples of $1/4$, which have been recently explained from the fractal branching architecture of organisms⁴. The challenge is to find body size scaling laws and their explanations for interactions between individuals and their environments¹⁹.

Home range, the area used by an animal in its daily and seasonal movements, integrates organism physiology, morphology and behaviour to determine patterns of space use, population density and interactions with other species^{5–7}. As a result, land managers and conservation biologists frequently use home range–body size scaling relationships to determine minimum reserve size and evaluate extinction risks of species⁸. Early studies proposed, and found in preliminary data, that if home range size is directly proportional to resource needs, then $b \cong 3/4$ (refs 9–11). Later studies^{12–14} with more data, however, found that, for all trophic groups combined, $b \cong 1$ and explained deviation in b from $3/4$ *post hoc* as resulting from larger animals encountering a lower resource density and thus causing home range to scale more steeply with size than expected from resource requirements¹⁵. An alternative explanation proposed that home range overlap was greater at larger body sizes, thus reducing resource availability per individual and increasing home range size above that expected from requirements^{5,16}. However, no specific mechanism was proposed for why resource density should be lower or home ranges overlap more for larger species. Lindstedt *et al.*¹⁷ argued that home range should be

proportional to metabolic requirements, which scale as $M^{3/4}$, multiplied by biological time, which scales as $M^{1/4}$, so home range should scale as M^1 . Holling⁶ first proposed that constraints on space use by animals could explain why home range should scale isometrically ($b = 1$): fractal habitat structure implies that smaller species exploit finer-grained features of the habitat than larger species and thus detect more resources. However, this theory did not explicitly address the issue of resource density and distribution. Most recently, Kelt and VanVuren¹⁸ proposed that home range does not exhibit a simple scaling relationship with size, as it changes nonlinearly according to size-dependent physiological constraints on animal's reproductive output and resource use. However, none of these hypotheses accounted for observed variation in exponents among trophic groups (for example, herbivores and carnivores¹⁵) and taxa (for example, birds and mammals^{12,13}). Thus, the evidence and theoretical explanations for a universal home range scaling law remain weak.

Here, we propose a new model to predict home range scaling with body size as a function of the spatial distribution and abundance of resources. This model provides the first theoretical synthetic explanation for why and how resource density scales with body size and the influence of resource density and distribution on home range scaling exponents. To survive, an individual must meet or exceed the energy it expends during maintenance, foraging, and reproduction B with resources gathered from the environment I such that $B \leq I$. To obtain resources, a forager instantaneously searches a volume of length w and dimension D as determined by whether its habitat is two- or three-dimensional. In two dimensions, for example, a forager would instantaneously search a square area with sides of length w . Over some time period t the forager samples v volumes per unit time, so the total volume $V(t)$ sampled over time t is tvw^D (refs 20, 21). An individual's home range results from movements over long time periods, for example, the resource renewal interval τ , or time required for a consumed resource to be renewed. Recent work²² shows that if τ is large then as $t \rightarrow \tau$, a random movement path will occupy a bounded volume, similar to a home range, H . Thus, as $t \rightarrow \tau$, $V(\tau) \approx H$.

Mounting evidence suggests that resource distributions are typically fractal-like: they exhibit statistically similar patterns over 2–3 orders of magnitude in scale of observation^{1,21,23,24}. A fractal resource distribution makes resource density scale dependent: different-sized species encounter different densities of resources in the same environment^{20,21}. To see this, consider that a forager maximizing its resource intake rate should minimize the number of sampling volumes needed for it to consume all resources in a landscape (Fig. 1). The resulting minimum number of sampling volumes will be spatially arranged to yield the maximum average

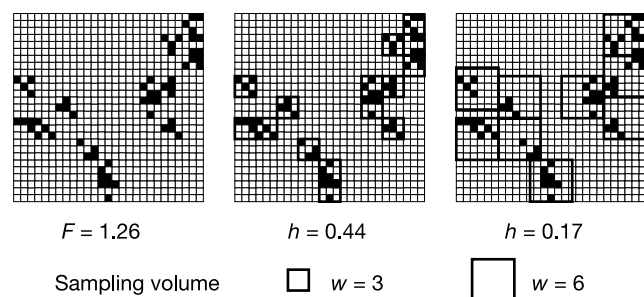


Figure 1 Hypothetical average density h of resources per sampling volume encountered by species with different sampling volumes of length w on a fractal distribution of resources (black cells, fractal dimension $F = 1.26$). The resource distribution was generated using random "Sierpinski gaskets"²⁵. Note that the species with the larger sampling volume encounters a lower average density of resources per sampling volume but requires fewer sampling volumes to incorporate all resources.

amount of resource per sampling volume^{1,23}. If so, the amount of resource per sampling volume will be, on average, $c_1 w^F$ (refs 1, 20, 21). The fractal dimension F describes the degree to which resources fill space and can range from 0, a single point, to 3, a solid cube^{1,20,21,25}. The constant c_1 describes resource density in the vicinity of other resources (clumping) as well as overall resource density^{1,20,21,25}. In general, both c_1 and F will increase as resource density increases (Fig. 2), but each can assume a range of values at most resource densities, depending on the spatial arrangement, and aggregation of the distribution²⁶. This property has a profound effect on the relationship of resource encounter rate with body size, and thus home range scaling: changes in resource density affect both the prefactor C_1 and the scaling exponents describing these relationships.

This scale-dependence of resource amount per sampling volume constrains the maximum resource density encountered by different sized foragers and thus total resource intake I and home range size (Fig. 1). Resource density r is the average amount of resources in a sampling volume $c_1 w^F$ divided by the sampling volume w^D , so $r = c_1 w^F / w^D = c_1 w^{F-D}$. Because F cannot exceed the dimension of the foraging habitat D over the same range of scales^{20,21}, $F \leq D$ and r must decline with increasing foraging scale w (refs 20, 27). Resource availability within the home range I within the resource renewal interval τ is the product of the home range H and r . Hence:

$$I = H c_1 w^{F-D} \quad (2)$$

This resource availability must be large enough to satisfy metabolic requirements B and this constrains the home range size to be at least:

$$H \geq \frac{B}{c_1 w^{F-D}} \quad (3)$$

This formula explicitly includes the effect of foraging scale on encountered resource abundance and thus home range.

To predict home range scaling relationships with body size, we can substitute well-known allometric functions for resource requirements: $B = c_2 M^{3/4}$ (refs 2–4), where c_2 is a taxon-specific

constant that includes species-specific resource demands over the resource renewal interval τ . We can also assume that the width of the search volume w is proportional to a physical distance, such as the forager's stride length³, which scales as $c_3 M^{1/3}$. These assumptions yield:

$$H = k M^{3/4 + D/3 - F/3} \quad (4)$$

where the prefactor $k = c_2 c_3^{D-F} / c_1$. The model predicts that there should not be a universal scaling exponent for home range size. The scaling exponent b is instead a function of the structure of the environment, as given by the dimension D , and the resource distribution as given by the dimension F . The exponent reflects the interaction between 1/4 power scaling of organism physiological rates, the structure of the environment, and the euclidean 1/3 power scaling of physical distances. The model predicts that, when $D = 2$, b will range from 3/4 when $F = 2$ to 17/12 as $F \rightarrow 0$. When $D = 3$, b will approach 21/12 as $F \rightarrow 0$. As expected if the minimum scale of resolution, $w = 1$, the coefficient k will increase with resource requirements (c_2) and movement rate (c_3) and decrease with greater resource density and/or clustering (c_1). Thus, for a given body mass, home range can vary widely depending on environmental con-

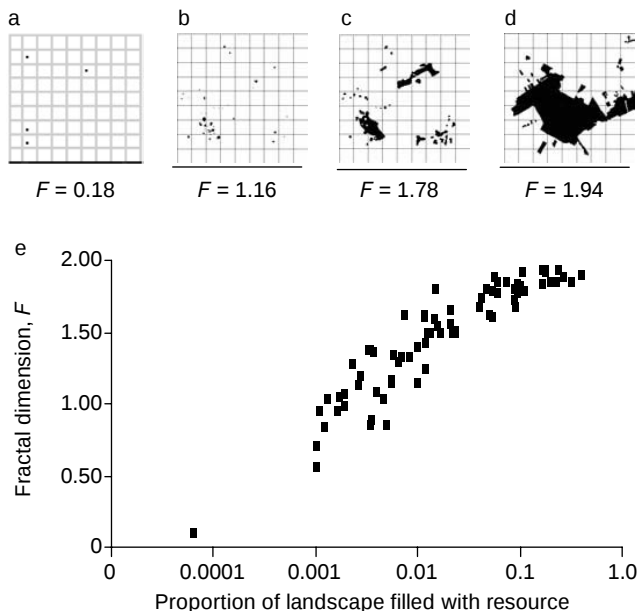


Figure 2 Relationship between density and estimated fractal dimension for different distributions in a two-dimensional environment. We illustrate this with real maps of some fractal distributions²⁶ of varying density, along a gradient of proportion of map filled (e). Although the fractal dimension varies at a given resource density, it increases with increasing density from almost 0 to 2 (a to d) over four orders of magnitude of proportional fill.

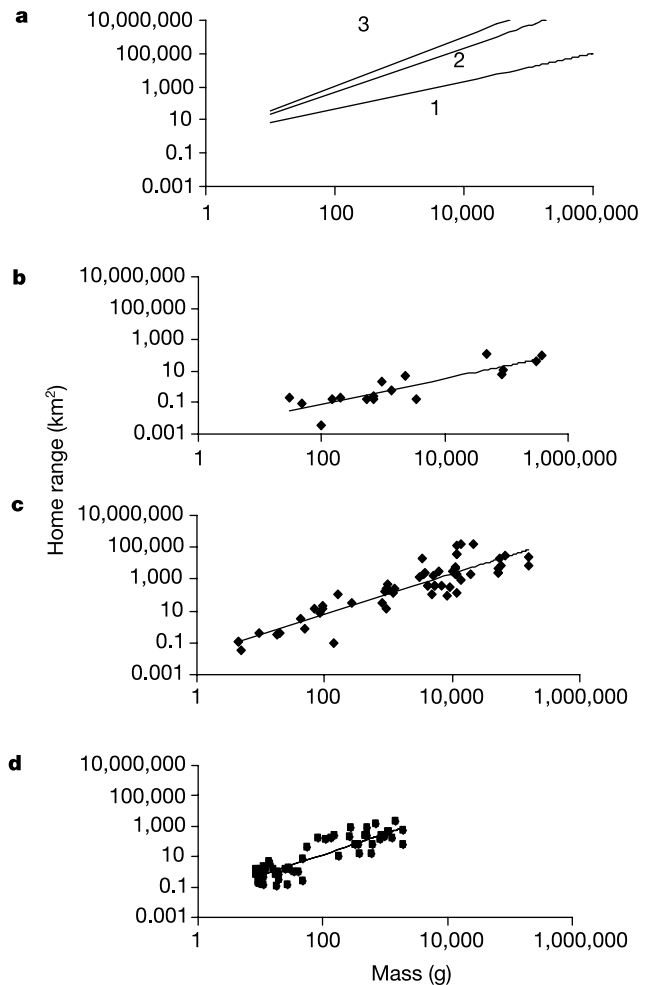


Figure 3 Comparison of model predictions and empirical estimates for home range-body mass scaling relationships (see also Table 1). **a**, Predicted scaling exponents b by trophic level: (1) terrestrial mammalian herbivores ($b \approx 5/6$); (2) terrestrial mammalian carnivores ($b \approx 4/3$); (3) terrestrial avian carnivores ($b \approx 3/2$). **b**, **c**, **d**, Observed relationships for terrestrial mammalian herbivores (**b**; $b \approx 0.83$, $R^2 = 0.74$) and terrestrial mammalian carnivores (**c**; $b \approx 1.21$, $R^2 = 0.80$) and terrestrial avian carnivores (**d**; $b \approx 1.37$, $R^2 = 0.78$; see Table 1).

Table 1 Predicted and observed home range scaling parameters

Foraging guild	Food resource	Expected F -values	Predicted b	Decimal equivalent	Observed b	s.e.	Observed Y_0	s.e.	N
Folivorous mammals*	Plant leaves	1.5–1.99	3/4–11/12	0.75–0.91	0.831	0.21	NA	NA	21
Folivorous mammals†	Plant leaves	1.5–1.99	3/4–11/12	0.75–0.91	0.802	0.11	0.0021	0.0003	16
Carnivorous mammals	Small vertebrates	0.5–1	15/12–17/12	1.08–1.25	1.213	0.11	0.0234	0.0053	32
Carnivorous birds	Small animals	0.5–1	17/12–19/12	1.41–1.58	1.374	0.09	0.0257	0.0031	58

Predicted home range scaling exponents (b) and coefficients (Y_0) compared to observed values from relationships of home range (km^2) versus body mass (in g) for herbivorous and carnivorous terrestrial mammals and carnivorous terrestrial birds^{5,6,9,13,15,28,29,30}.

*Regression results from Owen-Smith¹⁶; NA, individual data not available. Group home range divided by group size to estimate the area used by an individual.

†Home range divided by density to estimate the area used by an individual⁹.

Carnivorous mammals are non-social carnivores^{6,9,13,15}. Carnivorous birds are insectivorous and flesh-eating birds^{5,28,29}.

ditions, locomotory capacity and resource density.

Home range–body size relationships for carnivorous and herbivorous mammals and carnivorous birds strongly support the predictions of the model. Data from published sources^{5,6,9,13,15,28,29} were combined into one data set and an average mass and home range was recalculated for each species (see Methods). We estimated the range of likely fractal dimensions for plant leaves^{1,21} and for the prey of carnivores from published sources and data from Portal, Arizona³⁰. As demonstrated in Fig. 2, F varies predictably with resource density. Low-density resources in two dimensions such as small mammal prey should have F -values ranging from near 0 to 0.5. In three dimensions, small animal prey such as insects and small vertebrates should have F -values ranging from 0.5 to 1. Relatively high-density resources such as plant leaves should have F -values between 1.5 and 2. In our model, these estimates suggest that b for mammals feeding in a two-dimensional environment should range from 3/4 to 11/12 for herbivores, and from 13/12 to 15/12 for carnivores. Furthermore b should range from 17/12 to 19/12 for carnivorous birds feeding in three-dimensional environments. As shown in Table 1, these predictions correspond closely with the results of our analyses (Fig. 3).

The model has profound implications for ecologists and wildlife managers because it shows how resource spatial distribution, and thus habitat fragmentation, might affect home range. Home range scaling exponents larger than 3/4 suggest that larger species use much larger areas than would be expected on the basis of their resource requirements. Furthermore, habitat fragmentation that results in a loss of resource density and thus lower c_1 and F may greatly increase home range size, and this increase will be much greater for larger species. In a two-dimensional environment ($D = 2$), for example, the sensitivity of predicted home range size H (equation (4)) to changes in c_1 is $\partial H / \partial c_1 = -c_2 c_3^{-2-F} M^{17/12-F/3} / c_1^2$. This result implies that a decrease in c_1 will increase H and the increase will be greater for larger species, because $\partial H / \partial c_1$ is proportional to M . Likewise, the sensitivity of home range size to changes in the resource fractal dimension F is $\partial H / \partial F = -[\ln(c_1) + (1/3)\ln(M)]c_2 c_3^{-2-F} M^{17/12-F/3} / c_1$. This result implies that a decrease in F will also increase H and do so more strongly for larger species. Thus, habitat loss will require greater compensatory changes in home ranges of larger species, and lead to a steeper body size scaling relationship for home range. These predictions emphasize the high potential vulnerability of large carnivores to habitat fragmentation. Carnivores have naturally low-density prey, and thus low c_1 and F -values, and therefore little ability to compensate for reductions in c_1 and F . In contrast, herbivore species, whose plant resources tend to occur at much higher density and thus fractal dimension, can potentially expand their home ranges by 2–3 orders of magnitude to compensate for habitat fragmentation.

Our home range model, based on an assumption of fractal resource distributions, predicts that there is no universal scaling law for home range. Instead, the range of variation in scaling exponents, b , is constrained between a minimum of 3/4 and a maximum of 17/12 for organisms in two-dimensional environments, and a maximum of 21/12 in three-dimensional environ-

ments. The model unifies previous hypotheses of home range scaling by incorporating two fundamental properties of fractal-like environments: (1) the relationship between resource distribution F and resource density r (Fig. 2); and (2) the decline in resource density with increasing foraging scale w (Fig. 1). These properties provide a potential framework for understanding scaling relationships for other interactions between individuals and their environment^{20,21,23}. Here we have generated a framework for exploring important sources of variation in home range scaling. These include differences in scaling exponents among different trophic and taxonomic groups and the effects of habitat fragmentation. Clearly, there are many factors besides body size that influence home range size, such as population density, temperature and other factors that influence resource demand, that we have not included explicitly in our model. However, it seems unlikely that these factors would affect the body size scaling exponents for home range. The agreement of the scaling exponents predicted by our model with exponents of observed relationships suggests that resource distribution and density are critical factors affecting home range scaling relationships. □

Methods

Home range data for bird and mammal species were taken from published sources^{5,6,9,13,15,28,29}. In our analysis, we only included those species (1) for which we unambiguously knew the type of food resources, and (2) from diet guilds for which we could estimate ranges of fractal dimensions of resource distribution. Data for individual species were averaged to obtain a single home range (in km^2) and body mass (in g) value for each species. Data was log transformed and least squares regression was used to estimate the observed home range scaling exponent.

Received 27 November 2001; accepted 24 April 2002; doi:10.1038/nature00840.

1. Milne, B. T. in *Wildlife and Landscape Ecology: Effects of Pattern and Scale* (ed. Bissonette, J. A.) 32–69 (Springer, New York, 1997).
2. Peters, R. H. *The Ecological Implications of Body Size* (Cambridge Univ. Press, New York, 1983).
3. Calder, W. A. *Size, Function, and Life History* (Dover, Mineola, New York, 1996).
4. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
5. Damuth, J. Home range, home range overlap, and species energy use among herbivorous mammals. *Biol. J. Linn. Soc.* **15**, 185–193 (1981).
6. Holling, C. S. Cross-scale morphology, geometry, and dynamics of ecosystems. *Ecol. Monogr.* **62**, 447–502 (1992).
7. Reiss, M. Scaling of home range size: body size, metabolic needs and ecology. *Trends Ecol. Evol.* **3**, 85–88 (1988).
8. Belovsky, G. E. in *Viable Populations for Conservation* (ed. Soule, M. E.) (Cambridge Univ. Press, Cambridge, 1987).
9. McNab, B. K. Bioenergetics and the determination of home range size. *Am. Nat.* **97**, 133–140 (1963).
10. Swihart, R. K., Slade, N. A. & Bergstrom, B. J. Relating body size to the rate of home range use in mammals. *Ecology* **69**, 393–399 (1988).
11. Whiting, L. The body mass allometries as evolutionarily determined by the foraging of mobile organisms. *J. Theor. Biol.* **177**, 129–137 (1995).
12. Armstrong, J. T. Breeding home range in the nighthawk and other birds: its evolutionary and ecological significance. *Ecology* **46**, 619–629 (1965).
13. Schoener, T. W. Sizes of feeding territories among birds. *Ecology* **49**, 123–141 (1968).
14. Gompfer, M. E. & Gittleman, J. L. Home range scaling: Intraspecific and comparative trends. *Oecologia* **87**, 343–348 (1991).
15. Harestad, A. S. & Bunnell, F. L. Home range and body weight—A reevaluation. *Ecology* **60**, 389–482 (1979).
16. Owen-Smith, R. N. *Megaherbivores: The Influence of Very Large Body Size on Ecology* (Cambridge Univ. Press, New York, 1992).
17. Lindstedt, S. L., Miller, B. J., Buskirk, S. W. Home range, time and body size in mammals. *Ecology* **67**, 413–418 (1986).
18. Kelt, D. A. & VanVuren, D. Energetic constraints and the relationship between body size and home range area in mammals. *Ecology* **80**, 337–340 (1999).
19. Lawton, J. H. Are there general laws in ecology? *Oikos* **84**, 177–192 (1999).

20. Ritchie, M. E. Scale-dependent foraging and patch choice in fractal environments. *Evol. Ecol.* **12**, 309–330 (1998).
21. Ritchie, M. E. & Olff, H. Spatial scaling laws yield a synthetic theory of biodiversity. *Nature* **400**, 557–560 (2000).
22. Lawler, G., Schramm, O. & Werner, W. The dimension of the planar brownian frontier is 4/3. *Math. Res. Lett.* **8**, 13–25 (2001).
23. Milne, B. T. Spatial aggregation and neutral models in fractal landscapes. *Am. Nat.* **139**, 32–57 (1992).
24. Kunin, W. E. Extrapolating species abundance across spatial scales. *Science* **281**, 1513–1515 (1998).
25. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1982).
26. Olff, H. & Ritchie, M. E. Fragmented nature: consequences for biodiversity. *Landscape Urban Planning* **858**, 1–10 (2001).
27. Kotler, N. B. & Wiens, J. A. Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. *Oikos* **59**, 253–260 (1990).
28. Poole, A. & Gill, F. (eds) *The Birds of North America* (The Birds of North America, Philadelphia, 1996–2000).
29. Dunning, J. B. *CRC Handbook of Avian Body Masses* (CRC, Boca Raton, 1993).
30. Brown, J. H. in *Experimental Ecology* (eds Reserits, W. J. Jr & Bernardo, J.) 71–95 (Oxford Univ. Press, New York, 1998).

Acknowledgements

We thank J. Brown, B. Enquist, J. Damuth and G. Belovsky. Work on this model began during the Fractals in Biology Meeting at the Santa Fe Institute, New Mexico. J.P.H. is supported by an NSF Graduate Research Fellowship.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.P.H. (e-mail: jhaskell@cc.usu.edu).

The endogenous cannabinoid system controls extinction of aversive memories

Giovanni Marsicano[☆], Carsten T. Wotjak[†], Shahnaz C. Azad[‡], Tiziana Bisogno^{||}, Gerhard Rammes[‡], Maria Grazia Cascio^{||}, Heike Hermann^{*}, Jianrong Tang[†], Clementine Hofmann[¶], Walter Ziegler[‡], Vincenzo Di Marzo^{||} & Beat Lutz^{*}

^{*} Molecular Genetics of Behaviour, [†] Neuronal Plasticity/Mouse Behaviour and [‡] Neuropharmacology Group, Max Planck Institute of Psychiatry, Kraepelinstrasse 2-10, 80804 Munich, Germany

[§] Clinic of Anaesthesiology, Klinikum Grosshadern, Ludwig-Maximilians-University, Marchioninistrasse 15, 81377 Munich, Germany

^{||} Endocannabinoid Research Group, Institute of Biomolecular Chemistry, CNR, Via Campi Flegrei 34, 80078 Pozzuoli (Napoli), Italy

[¶] Institute of Mammalian Genetics, GSF National Research Center for Environment and Health, Ingolstädter Landstrasse 1, 85764 Neuherberg-Munich, Germany

[☆] These authors contributed equally to this work

Acquisition and storage of aversive memories is one of the basic principles of central nervous systems throughout the animal kingdom¹. In the absence of reinforcement, the resulting behavioural response will gradually diminish to be finally extinct. Despite the importance of extinction², its cellular mechanisms are largely unknown. The cannabinoid receptor 1 (CB1)³ and endocannabinoids⁴ are present in memory-related brain areas^{5,6} and modulate memory^{7,8}. Here we show that the endogenous cannabinoid system has a central function in extinction of aversive memories. CB1-deficient mice showed strongly impaired short-term and long-term extinction in auditory fear-conditioning tests, with unaffected memory acquisition and consolidation.

Treatment of wild-type mice with the CB1 antagonist SR141716A mimicked the phenotype of CB1-deficient mice, revealing that CB1 is required at the moment of memory extinction. Consistently, tone presentation during extinction trials resulted in elevated levels of endocannabinoids in the basolateral amygdala complex, a region known to control extinction of aversive memories⁹. In the basolateral amygdala, endocannabinoids and CB1 were crucially involved in long-term depression of GABA (γ -aminobutyric acid)-mediated inhibitory currents. We propose that endocannabinoids facilitate extinction of aversive memories through their selective inhibitory effects on local inhibitory networks in the amygdala.

To study the involvement of the endogenous cannabinoid system in memory processing, we generated CB1-deficient mice (CB1^{-/-}; see Supplementary Information). CB1^{-/-} mice and CB1^{+/+} littermates were tested in auditory fear conditioning, which is highly dependent on the amygdala¹ and enables the dissection of different phases of memory formation, including acquisition, consolidation and extinction. Mice were trained to associate a tone with a foot-shock (conditioning). After conditioning, animals froze when

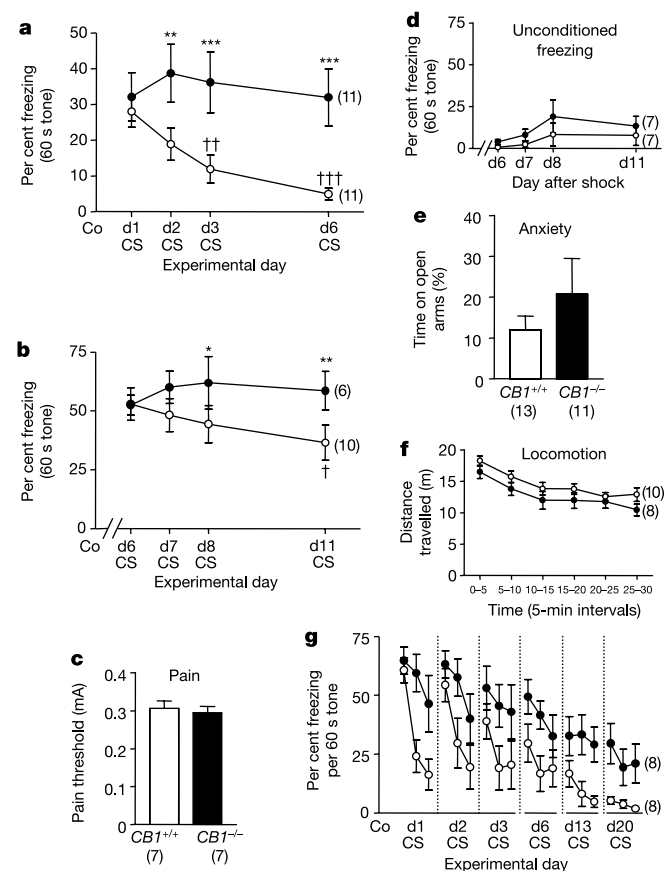


Figure 1 Impaired extinction of aversive memory in an auditory fear-conditioning task of CB1^{-/-} mice (filled circles) as compared to their CB1^{+/+} littermates (open circles). **a, b**, After conditioning (Co) animals were repeatedly exposed to 60 s tones (conditioned stimulus, CS) starting 24 h after conditioning (**a**) (d1) or after a 6-day consolidation period (**b**) (d6). **c–f**, CB1^{-/-} and CB1^{+/+} mice did not differ in their sensory-motor abilities, as assessed by sensitivity to rising electric foot-shock (**c**), unspecific freezing to a tone after shock application (**d**), anxiety-related behaviour on the elevated plus maze (**e**) and horizontal locomotion in an open field (**f**). **g**, CB1^{-/-} mice showed memory extinction in response to a stronger extinction protocol (3 min tones until day 20; analysed in 60-s intervals), but still froze more than CB1^{+/+} controls. Means $\pm$ s.e.m. are shown; number of animals are indicated in parentheses. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$ (compared with CB1^{+/+}); dagger, $P < 0.05$; double dagger, $P < 0.01$; triple dagger, $P < 0.001$ (compared with day 1).

[#] Present address: Molecular Neurogenetics Group, Max Planck Institute of Psychiatry, Kraepelinstrasse 2-10, 80804 Munich, Germany.

re-exposed to the tone. This response served as an indicator of aversive memory, and is gradually extinguished on repeated tone presentations. As the amygdala has a crucial role for extinction of aversive memories^{9,10}, we studied amygdala-dependent memory performance in the absence of possible confounding influences of the hippocampus by re-exposing the mice to the tone in an environment different from the conditioning context¹. In this environment, neither *CB1*^{-/-} nor *CB1*^{+/+} mice showed freezing without tone presentation 24 h after conditioning (data not shown). During the subsequent tone presentation, however, animals of both groups showed the same amount of freezing (Fig. 1a; d1, $P > 0.05$), pointing to an equally successful tone-foot-shock association. On repeated exposure to the tone, however, *CB1*^{+/+} and *CB1*^{-/-} mice differed significantly in their freezing behaviour (genotype: $F_{1,20} = 5.81$, $P < 0.05$; genotype $\times$ day interaction: $F_{3,60} = 4.86$, $P < 0.005$; Fig. 1a). In fact, *CB1*^{+/+} mice ($F_{3,10} = 9.70$, $P < 0.0005$), but not *CB1*^{-/-} ($F_{3,10} = 0.94$, $P = 0.433$), showed extinction of freezing.

The identical behavioural performance of the two genotypes on day 1 indicates that acquisition and early consolidation processes do not involve CB1. However, it is possible that memory consolidation processes were not completed 24 h after conditioning, leaving open a potential involvement of CB1 in later phases of memory consolidation. To test this hypothesis, new groups of animals remained undisturbed after conditioning for 6 days, and mice from these groups were then exposed to the 60-s tones (Fig. 1b). Again, *CB1*^{-/-} and *CB1*^{+/+} mice did not differ in their initial freezing response, but behaved in a significantly different way in the course of repeated tone presentations (genotype $\times$ day interaction: $F_{3,42} = 3.03$, $P < 0.05$). Whereas *CB1*^{+/+} mice showed a decrease in freezing behaviour until day 11 ($F_{3,27} = 3.73$, $P < 0.05$), *CB1*^{-/-} mice failed to extinguish the freezing response ($F_{3,15} = 1.03$, $P = 0.404$). A more detailed analysis of the freezing response in 20-s intervals confirmed the difference in extinction (genotype $\times$ 20-s bin interaction: $F_{11,154} = 2.60$, $P < 0.005$; Supplementary Information). These differences were due to altered short-term and long-term extinction in *CB1*^{-/-} mice but not to increased spontaneous recovery of the freezing response (genotype:

$F_{1,14} = 0.18$, $P = 0.675$; genotype $\times$ day interaction: $F_{2,28} = 1.61$, $P = 0.217$; Supplementary Information).

We next analysed whether the differences in memory extinction between the two genotypes could be attributed to alterations in sensory-motor abilities of *CB1*^{-/-} mice, as cannabinoids are known to influence pain perception, emotionality and locomotion^{4,11,12}. However, mice of either genotype showed the same pain sensitivity to a rising electric foot-shock defined as the shock intensity at which mice showed first signs of discomfort, that is, jumping and/or vocalization (Fig. 1c). Moreover, if the same animals were repeatedly exposed to the tone, there were no significant differences in freezing behaviour between the genotypes (genotype: $F_{1,12} = 1.61$, $P = 0.228$; genotype $\times$ day interaction: $F_{3,36} = 0.225$, $P = 0.878$; Fig. 1d), indicating that CB1 deficiency does not affect foot-shock-induced behavioural sensitization or unconditioned freezing to the tone. Anxiety-related behaviour was analysed on an elevated plus maze. Animals of either genotype spent the same relative time on open arms of the maze ($P > 0.05$, t -test and U -test; Fig. 1e), and made the same relative number of entries into open arms (*CB1*^{+/+}: $22.0 \pm 4.0\%$; *CB1*^{-/-}: $21.1 \pm 7.6\%$, $P > 0.05$, t -test and U -test). In contrast, *CB1*^{-/-} mice showed reduced exploratory activity (number of closed-arm entries: 11.6 ± 1.1 in *CB1*^{+/+} mice compared with 6.5 ± 1.2 in *CB1*^{-/-} mice, $P < 0.01$, t -test). However, in an open-field locomotor activity test, no significant differences were found, including horizontal (Fig. 1f) and vertical locomotion, resting time, and time spent close to the walls of the box (data not shown).

The failure of *CB1*^{-/-} mice to diminish their freezing response during a limited number of 60-s tone presentations (Fig. 1a, b) raises the question as to whether *CB1*^{-/-} mice are able to extinguish aversive memories at all. Thus, conditioned *CB1*^{-/-} and *CB1*^{+/+} mice were exposed to a stronger extinction protocol (3 min tone, six exposures; Fig. 1g). Both *CB1*^{+/+} ($F_{17,119} = 15.01$, $P < 0.000001$) and *CB1*^{-/-} mice ($F_{17,119} = 7.59$, $P < 0.000001$) extinguished their freezing response over the course of repeated tone presentations. Nevertheless, extinction was still more pronounced in *CB1*^{+/+} as compared with *CB1*^{-/-} mice (genotype: $F_{1,14} = 5.30$, $P < 0.05$). Notably, the most marked differences between *CB1*^{-/-}

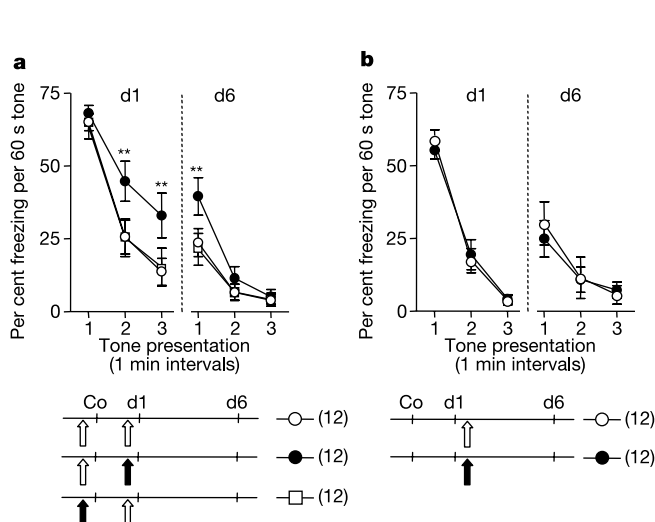


Figure 2 CB1 antagonist SR141716A impairs short-term and long-term extinction, but not acquisition and consolidation of aversive memories. **a**, Mice were treated with SR141716A (filled arrows) or vehicle (open arrows) 20 min before conditioning (Co) and the first extinction trial (d1; 3 min tone). **b**, Mice were treated with SR141716A or vehicle 10 min after the first extinction trial, as indicated. Freezing was analysed in 60-s intervals. Means $\pm$ s.e.m. are shown; number of animals are shown in parentheses. Double asterisk, $P < 0.01$ (compared with the two other groups).

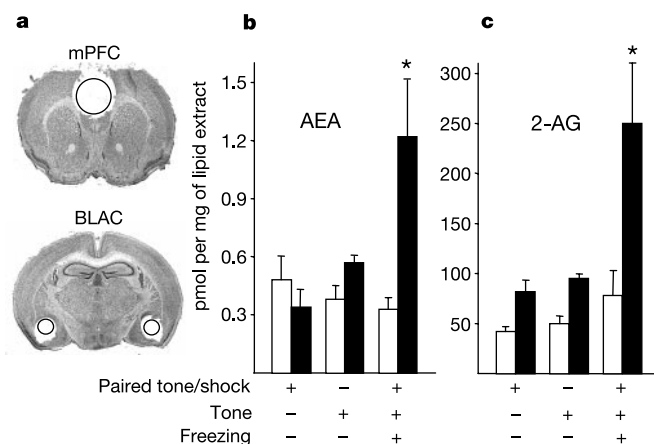


Figure 3 Re-exposure to the tone 24 h after conditioning causes increased endocannabinoid levels in the basolateral amygdala complex (BLAC) but not the medial prefrontal cortex (mPFC) of C57BL/6J mice. **a**, Micrographs of coronal brain sections showing representative examples of the dissected mPFC and BLAC. Circles indicate the size and positioning of tissue sampling. **b**, **c**, Anandamide (AEA) and 2-arachidonoylglycerol (2-AG) levels of the three experimental groups (see text), which differed in conditioning procedure, re-exposure to the tone and resulting freezing response to the tone. Means $\pm$ s.e.m. are shown ($n = 4$ per group, 5 mice per n). Open bars, mPFC; filled bars, BLAC. Asterisk, $P < 0.05$ (compared with BLAC of the other groups).

and $CB1^{+/+}$ mice were observed during acute tone presentation (short-term extinction). Therefore, $CB1^{-/-}$ mice might be primarily impaired in short-term extinction, with a resulting impairment in long-term extinction, assessed in the course of the subsequent extinction trials. Accordingly, spontaneous recovery was not different between the genotypes (genotype: $F_{1,14} = 1.73$, $P = 0.208$; genotype $\times$ day interaction: $F_{4,56} = 1.19$, $P = 0.323$; Supplementary Information).

Our behavioural data clearly indicate an involvement of the endogenous cannabinoid system in extinction of aversive memories. However, the life-long absence of CB1 could result in developmental defects leading to the phenotype observed. It, furthermore, precludes any temporal dissection of the involvement of the endogenous cannabinoid system in different stages of memory formation. Thus, we treated wild-type C57BL/6J mice with the CB1 antagonist SR141716A (ref. 13), either before conditioning, or before the first extinction trial. Systemic application of SR141716A 20 min before the first extinction trial impaired both short-term and

long-term extinction of the freezing response as compared with both vehicle-treated controls and animals treated with SR141716A before conditioning (treatment $\times$ time interaction: $F_{10,160} = 2.72$, $P < 0.005$), with no difference between the two latter treatments and with a similar performance of all three groups in the beginning of the first extinction trial (Fig. 2a). These data largely confirm the phenotype of $CB1^{-/-}$ mice (Fig. 1a, b, g), indicating that endocannabinoids have only a negligible function in memory acquisition, consolidation and recall (indicated by the similar performance at the beginning of the first extinction trial), but selectively interfere with extinction of the freezing response to the tone. Mice treated with SR141716A before the first extinction trial showed an attenuated extinction of freezing not only during the first tone presentation (short-term extinction) but also in the absence of pharmacological treatment during the first 60 s of tone presentation at day 6 (long-term extinction). Spontaneous recovery of the behavioural performance from the end of the first (day 1) to the beginning of the second tone presentation session (day 6) was not different among the three groups ($F_{2,34} = 0.29$, $P = 0.744$; Supplementary Information). Together, these findings support the idea that CB1 might be particularly important for the extinction of acute responses to the tone (short-term extinction), which, in turn, relates to behavioural extinction over repeated tone presentations (long-term extinction), without affecting spontaneous recovery of the behavioural performance. Accordingly, the CB1 antagonist had to be present at the time of tone presentation (that is, during aversive memory recall) in order to interfere with memory extinction, as SR141716A failed to affect extinction if administered immediately at the end of the extinction trial (data not shown) or 10 min later (Fig. 2b).

These observations, together with the pharmacokinetics of SR141716A (ref. 14), led us to assume that presentation of the tone during the extinction trial causes an instantaneous rise in endocannabinoid levels. To confirm this assumption, we measured in C57BL/6J mice levels of the two major endocannabinoids, anandamide (AEA) and 2-arachidonoylglycerol (2-AG), in brain punches of the medial prefrontal cortex (mPFC) and the basolateral amygdala complex (BLAC), both of which are thought to have central roles in extinction of aversive memories^{9,15}. In those animals forming an association between tone and foot-shock, levels of AEA and 2-AG were significantly higher in the BLAC at the end of tone presentation of the extinction trial on day 1, as compared with animals with unpaired tone and foot-shock presentation on the previous day and with animals with paired tone and foot-shock presentation but no re-exposure to the tone (Fig. 3). There were no significant differences in levels of AEA and 2-AG in the mPFC, suggesting a specific involvement of endocannabinoids in extinction processes within the BLAC. Data of the two control groups indicate that both a successful tone-foot-shock association and re-exposure to the tone are required to trigger the acute increase of endocannabinoid levels.

If the endogenous cannabinoid system is activated during tone presentation, how exactly does it facilitate memory extinction? To answer this question, we performed a series of electrophysiological experiments in the BLAC of brain slices from $CB1^{-/-}$ and $CB1^{+/+}$ mice. Basic electrical properties were similar in $CB1^{-/-}$ and $CB1^{+/+}$ littermates, including input resistance and resting membrane potential (data not shown). High-frequency stimulation (HFS) in the lateral amygdala close to the external capsule induced long-term potentiation (LTP) in the basolateral amygdala of both genotypes (Fig. 4a). This effect was significantly more pronounced in $CB1^{-/-}$ than in $CB1^{+/+}$ mice (potentiation of population spike amplitude to $147 \pm 11\%$ in $CB1^{-/-}$ compared with $117 \pm 8\%$ in $CB1^{+/+}$ mice, $n = 9$, $P < 0.05$). However, we failed to affect basal synaptic transmission and LTP induction in wild-type slices superfused with SR141716A ($5 \mu\text{M}$; data not shown). This indicates that the enhanced LTP in $CB1^{-/-}$ mice might reflect long-term develop-

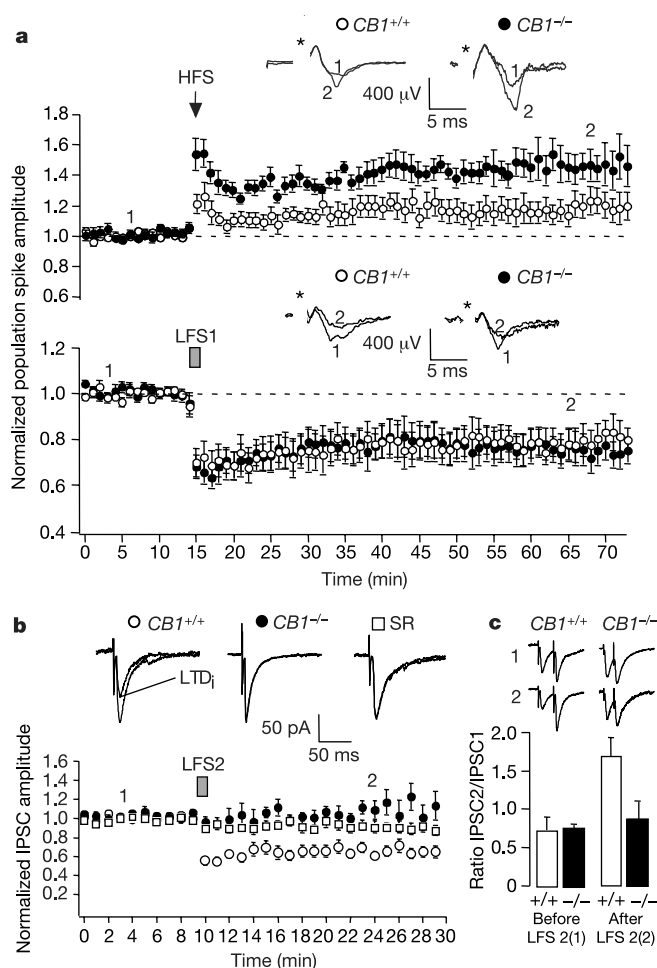


Figure 4 Endogenous cannabinoid system and synaptic plasticity in the basolateral amygdala. **a**, LTP (top) and LTD (bottom) in slices from $CB1^{+/+}$ and $CB1^{-/-}$ mice, induced by high-frequency stimulation (HFS) and low-frequency stimulation (LFS 1), respectively. Asterisks indicate stimulus artefacts. **b**, Long-term depression of IPSCs (LTD) requires CB1 activation. In principal neurons of slices of $CB1^{+/+}$ mice, low-frequency stimulation (LFS 2) induced a reduction of the amplitudes of isolated IPSCs. Slices of $CB1^{+/+}$ mice pre-incubated in SR141716A (SR) showed no LTD. LFS 2 had no effect in $CB1^{-/-}$ mice. **c**, LTD₁ was accompanied by increased PPF, which was absent in $CB1^{-/-}$ mice. Insets show representative traces before and after HFS or LFS (1, 2, respectively). Means $\pm$ s.e.m. are shown.

mental adaptations to life-long absence of CB1, and cannot be easily attributed to the lack of CB1 during LTP induction. Low-frequency stimulation with 900 pulses at 1 Hz (LFS 1) of the same pathway induced a persistent decrease in excitatory synaptic transmission (long-term depression, LTD) in both *CB1*^{-/-} and *CB1*^{+/+} mice with no difference between genotypes (depression of population spike amplitude to $75 \pm 7\%$ in *CB1*^{-/-} compared with $80 \pm 7\%$ in *CB1*^{+/+} mice, $n = 9$, $P > 0.05$; Fig. 4a).

As several recent studies indicate an involvement of CB1 in GABA-mediated synaptic transmission in hippocampus^{16,17} and amygdala⁶, we next looked for possible differences in this process within the basolateral amygdala of *CB1*^{-/-} and *CB1*^{+/+} mice. Low-frequency stimulation with 100 pulses at 1 Hz (LFS 2) of the lateral amygdala close to the external capsule induced a significant suppression of isolated GABA_A receptor-mediated inhibitory post-synaptic currents (IPSCs) in principal neurons of the basolateral amygdala of *CB1*^{+/+} mice. This suppression lasted for more than 20 min (hereafter called long-term depression of IPSCs, LTD_i, to $66.7 \pm 5.4\%$, $n = 8$, $P < 0.05$; Fig. 4b). Importantly, LTD_i was blocked in *CB1*^{+/+} mice by SR141716A (5 μ M; Fig. 4b), showing an acute involvement of the endocannabinoid system in the development of LTD_i. The involvement of CB1 in LTD_i was confirmed in *CB1*^{-/-} mice in which LTD_i was completely abolished (to $110.1 \pm 13.8\%$, $n = 8$, $P < 0.01$ compared with *CB1*^{+/+}; Fig. 4b). Consistent with previous reports^{16,17}, suppression of GABA-mediated synaptic transmission also increased paired-pulse facilitation (PPF) in *CB1*^{+/+} ($P < 0.05$) but not in *CB1*^{-/-} mice (Fig. 4c), indicating a local CB1-dependent decrease in GABA release from axon terminals in *CB1*^{+/+} slices.

Extinction of aversive memories is thought to be an active mnemonic process². As a new memory, it shares several attributes with other steps of memory formation^{9,10,18}; however, there is increasing evidence that some cellular pathways are specifically involved in extinction, but not in acquisition or consolidation of fear memories^{15,19,20}. We demonstrated a specific involvement of CB1-mediated neurotransmission in extinction of aversive memories. In principle, the enhanced excitatory synaptic plasticity in *CB1*^{-/-} mice (LTP; Fig. 4a) might explain the prolonged maintenance of aversive memories observed in these animals (Fig. 1a, b, g). However, an enhanced LTP is expected to coincide with an increased initial freezing response in the first extinction trial²¹, which was not observed in *CB1*^{-/-} mice. Accordingly, acute blockade of CB1 by a selective antagonist failed to affect LTP induction as well as acquisition and consolidation of the aversive memory. In contrast, the same approach revealed a significant involvement of CB1 in extinction (Fig. 2a). Tone-induced recall of the aversive memory was accompanied by an activation of the endocannabinoid system within the BLAC (Fig. 3), which possibly leads to a decrease of GABA-mediated transmission in a CB1-dependent manner (LTD_i; Fig. 4b, c).

The role of GABA-mediated transmission for extinction is, however, controversial^{22,23}. Within the amygdala, CB1 immunoreactivity was detected in a distinct subset of GABA-containing interneurons of the BLAC⁶ (one of the sites where aversive memories might be formed and stored²⁴), but not in the central nucleus of the amygdala⁶ (the principal output site of the amygdala¹). Taking into consideration that principal neurons of the BLAC and neurons of the central nucleus of the amygdala might be inversely correlated in their activities^{25,26}, we propose that the CB1-mediated decrease of activity of local inhibitory networks within the BLAC leads to a disinhibition of principal neurons and finally to extinction of the freezing response. The selective and locally restricted inhibition of GABA-mediated transmission might not be easily reproduced by systemic administration of GABA-interfering drugs^{22,23}. Thus, future studies will have to confine such treatments to the BLAC to validate that CB1-mediated inhibition of GABA-mediated transmission is indeed crucially involved in the extinction of

aversive memories mediated by CB1. It remains to be shown whether CB1 is not only involved in extinction of aversive memories but also in adaptation to aversive situations in general and/or in extinction of memories, independently from their emotional value.

Overall, our findings suggest that the endogenous cannabinoid system could represent a therapeutic target for the treatment of diseases associated with inappropriate retention of aversive memories or inadequate responses to aversive situations, such as post-traumatic stress disorders², phobias, and certain forms of chronic pain¹¹. □

Methods

Animals

Adult male C57BL/6J OlaHsd mice (6–8 weeks; Harlan–Winkelmann) and male *CB1*^{-/-} and *CB1*^{+/+} littermates (10–16 weeks; see Supplementary Information) were housed individually with an inverse 12/12 h light/dark cycle (lights off at 8:00) for at least 2 weeks before starting the experiments.

Behavioural studies

Experimental procedures were approved by the Committee on Animal Health and Care of local Government. Experiments were performed between 9:00 and 14:00. Animal's behaviour was analysed in a blind fashion with regards to genotype and drug treatment. Data were analysed by analysis of variance (ANOVA) followed by Fisher's least significant difference test for planned comparisons, Mann-Whitney *U*-test or unpaired Student's *t*-test. A *P*-value of <0.05 was considered statistically significant. Experimental procedures for pain threshold and unconditioned freezing, elevated plus maze and open field are described in Supplementary Information.

Fear conditioning

For conditioning, animals were placed into conditioning chambers (MED Associates). After 3 min, a 20-s tone (9 kHz, 80 dB) was presented that co-terminated with a 2-s electric foot-shock (0.7 mA). In pharmacological experiments animals received a 1-s shock to avoid ceiling effects in the freezing response due to the combination of foot-shock and injection stress. Animals were returned to their home cages 60 s after shock application. At the given time points after conditioning, animals were placed into transparent plexiglas cylinders that differed from the conditioning context, and a 60-s or 180-s tone was presented 3 min later (extinction trials). Animals were returned to their home cages after another 60 s. Mice were experimentally naive except for the stronger extinction protocol, where they had been tested on the elevated plus maze 5 days before. Freezing behaviour (defined as the absence of all movements except for respiration) was quantified from videotapes by trained observers that were blind to genotype and drug treatment, and data were normalized to the respective observation periods.

Pharmacological treatment

SR141716A (NIMH Chemical Synthesis and Drug Supply Program) was dissolved in vehicle solution (1 drop of Tween-80 in 3 ml 2.5% dimethylsulphoxide in saline). SR141716A (3 mg per kg body weight) and vehicle were injected subcutaneously at 20 ml per kg body weight under light isoflurane anaesthesia.

Measurement of endocannabinoids

C57BL/6J OlaHsd mice were randomly assigned to three groups ($n = 20$ each). On the conditioning day, two groups were conditioned as described before (paired). The remaining group received the foot-shock first and a 20 s tone 3 min later (unpaired). On the next day, all animals were placed into the cylinders, but only one of the paired groups and the unpaired group were exposed to a 3-min tone. Immediately after the end of the tone (or equivalent time in cylinder), animals were killed, brains were quickly removed and snap-frozen in isopentane/dry ice. mPFC and BLAC were punched from the frozen brain using a cryocut and cylindric brain punchers (Fine Science Tools, internal diameter 2.0 mm and 0.8 mm, respectively). Length of punches was approximately 1.6 mm for mPFC (start: bregma +2.8 mm²⁷) and 1.2 mm for BLAC (start: bregma -1.0 mm²⁷). Brain tissue of mPFC and bilateral BLAC, respectively, of 5 mice was pooled to obtain a single data point. Tissues (10–15 mg per data point) were dounce-homogenized with chloroform/methanol/Tris-HCl 50 mM, pH 7.4 (1/1/1 by volume) containing 5 pmol of octa-deuterated (d₈)-anandamide and 50 pmol of d₈-2-arachidonoylglycerol (Cayman Chemicals) as internal standards. Lipid-containing organic phase was dried down, weighed and pre-purified by open-bed chromatography on silica gel, and analysed by liquid chromatography-atmospheric pressure chemical ionization-mass spectrometry (LC-APCI-MS) using a Shimadzu high-performance liquid chromatography (HPLC) apparatus (LC-10ADVP) coupled to a Shimadzu quadrupole mass spectrometer (LCMS-2010) via a Shimadzu APCI interface. Mass spectrometry analyses were carried out in the selected ion-monitoring (SIM) mode as described previously²⁸. Temperature of the APCI source was 400 °C; HPLC column was a Phenomenex (5 μ m, 150 $\times$ 4.5 mm) reverse phase column, eluted as described²⁸. Anandamide (retention time of 14.5 min) and 2-AG (retention time of 17.0 min) quasi-molecular ions were quantified by isotope dilution with the above-mentioned deuterated standards²⁸ and their amounts in pmols normalized per mg of lipid extract. Data were statistically evaluated by ANOVA.

Electrophysiology

Brain slices were prepared essentially as described²⁹. IPSCs and population spikes were evoked by square pulse stimuli (0.066 Hz, 5–12 mA, 200 μ s) delivered by means of bipolar tungsten electrodes positioned within the lateral amygdala close to the external capsule. Population spikes were recorded in the basolateral amygdala close to lateral amygdala using glass microelectrodes (2–3 M Ω) filled with artificial cerebrospinal fluid (ACSF)²⁹. HFS (five trains at 100 Hz for 1 s, 10-s interstimulus interval) was applied to induce LTP, and LFS1 (900 pulses at 1 Hz) was applied to induce LTD. Whole-cell GABA-mediated currents were isolated by adding NBQX (0.005 mM) and D(-)-2-amino-5-phosphopentanoic acid (AP5; 0.05 mM) to ACSF (bubbled with 95% O₂/5% CO₂; pH 7.3), and were recorded from visually identified somata of principal neurons of the basolateral amygdala³⁰ by glass electrodes (4.5–5 M Ω)¹⁶ containing (in mM): Mg-ATP 2, CsCH₃SO₃ 100, CsCl 60, EGTA 0.2, HEPES 10, MgCl₂ 1, QX314 5 and Na₃GTP 0.3 (pH 7.3). Patch clamp experiments were performed at $24 \pm 1^\circ\text{C}$ at a holding potential of -70 mV. LTD₂ was induced by 100 stimuli at 1 Hz (LFS 2). PPF was induced as described³⁰. Data are expressed as means $\pm$ s.e.m. We tested significance using the Student's *t*-test.

Received 24 December 2001; accepted 16 April 2002; doi:10.1038/nature00839.

- LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
- Davis, M., Falls, W. A. & Gewirtz, J. in *Contemporary Issues in Modeling Psychopathology* (eds Myslobodsky, M. S. & Weiner, I.) 113–141 (Kluwer Academic, Norwell, 2000).
- Matsuda, L. A., Lolait, S. J., Brownstein, M. J., Young, A. C. & Bonner, T. I. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature* **346**, 561–564 (1990).
- Di Marzo, V., Melck, D., Bisogno, T. & De Petrocellis, L. Endocannabinoids: endogenous cannabinoid receptor ligands with neuromodulatory action. *Trends Neurosci.* **21**, 521–528 (1998).
- Marsicano, G. & Lutz, B. Expression of the cannabinoid receptor CB1 in distinct neuronal subpopulations in the adult mouse forebrain. *Eur. J. Neurosci.* **11**, 4213–4225 (1999).
- Katona, I. *et al.* Distribution of CB1 cannabinoid receptors in the amygdala and their role in the control of GABAergic transmission. *J. Neurosci.* **21**, 9506–9518 (2001).
- Hampson, R. E. & Deadwyler, S. A. Role of cannabinoid receptors in memory storage. *Neurobiol. Dis.* **5**, 474–482 (1998).
- Reibaud, M. *et al.* Enhancement of memory in cannabinoid CB1 receptor knock-out mice. *Eur. J. Pharmacol.* **379**, R1–R2 (1999).
- Falls, W. A., Miserendino, M. J. & Davis, M. Extinction of fear-potentiated startle: blockade by infusion of an NMDA antagonist into the amygdala. *J. Neurosci.* **12**, 854–863 (1992).
- Lu, K. T., Walker, D. L. & Davis, M. Mitogen-activated protein kinase cascade in the basolateral nucleus of amygdala is involved in extinction of fear-potentiated startle. *J. Neurosci.* **21**, RC162 (2001).
- Pertwee, R. G. Cannabinoid receptors and pain. *Prog. Neurobiol.* **63**, 569–611 (2001).
- Rodriguez de Fonseca, F., Del Arco, I., Martin-Calderon, J. L., Gorriti, M. A. & Navarro, M. Role of the endogenous cannabinoid system in the regulation of motor activity. *Neurobiol. Dis.* **5**, 483–501 (1998).
- Rinaldi-Carmona, M. *et al.* SR141716A, a potent and selective antagonist of the brain cannabinoid receptor. *FEBS Lett.* **350**, 240–244 (1994).
- Petit, F., Jeantaud, B., Bertrand, P. & Imperato, A. Cannabinoid penetration into mouse brain as determined by *ex vivo* binding. *Eur. J. Pharmacol.* **374**, 417–421 (1999).
- Morgan, M. A., Romanski, L. M. & LeDoux, J. E. Extinction of emotional learning: contribution of medial prefrontal cortex. *Neurosci. Lett.* **163**, 109–113 (1993).
- Wilson, R. I. & Nicoll, R. A. Endogenous cannabinoids mediate retrograde signalling at hippocampal synapses. *Nature* **410**, 588–592 (2001).
- Ohno-Shosaku, T., Maejima, T. & Kano, M. Endogenous cannabinoids mediate retrograde signals from depolarized postsynaptic neurons to presynaptic terminals. *Neuron* **29**, 729–738 (2001).
- Santini, E., Muller, R. U. & Quirk, G. J. Consolidation of extinction learning involves transfer from NMDA-independent to NMDA-dependent memory. *J. Neurosci.* **21**, 9009–9017 (2001).
- El Ghundi, M., O'Dowd, B. F. & George, S. R. Prolonged fear responses in mice lacking dopamine D1 receptor. *Brain Res.* **892**, 86–93 (2001).
- Quirk, G. J., Russo, G. K., Barron, J. L. & Lebron, K. The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *J. Neurosci.* **20**, 6225–6231 (2000).
- Tang, Y. P. *et al.* Genetic enhancement of learning and memory in mice. *Nature* **401**, 63–69 (1999).
- Harris, J. A. & Westbrook, R. F. Evidence that GABA transmission mediates context-specific extinction of learned fear. *Psychopharmacology (Berl.)* **140**, 105–115 (1998).
- McGaugh, J. L., Castellano, C. & Brioni, J. Picrotoxin enhances latent extinction of conditioned fear. *Behav. Neurosci.* **104**, 264–267 (1990).
- Medina, J. F., Repa, C. J., Mauk, M. D. & LeDoux, J. E. Parallels between cerebellum- and amygdala-dependent conditioning. *Nature Rev. Neurosci.* **3**, 122–131 (2002).
- Collins, D. R. & Paré, D. Reciprocal changes in the firing probability of lateral and central medial amygdala neurons. *J. Neurosci.* **19**, 836–844 (1999).
- Royer, S., Martina, M. & Paré, D. An inhibitory interface gates impulse traffic between the input and output stations of the amygdala. *J. Neurosci.* **19**, 10575–10583 (1999).
- Franklin, K. B. J. & Paxinos, G. *The Mouse Brain* (Academic, San Diego, 1997).
- Di Marzo, V. *et al.* Levels, metabolism, and pharmacological activity of anandamide in CB(1) cannabinoid receptor knockout mice: evidence for non-CB(1), non-CB(2) receptor-mediated actions of anandamide in mouse brain. *J. Neurochem.* **75**, 2434–2444 (2000).
- Rammes, G. *et al.* Synaptic plasticity in the basolateral amygdala in transgenic mice expressing dominant-negative cAMP response element-binding protein (CREB) in forebrain. *Eur. J. Neurosci.* **12**, 2534–2546 (2000).
- Rammes, G., Eder, M., Dodt, H.-U., Kochs, E. & Zieglgänsberger, W. Long-term depression in the basolateral amygdala of the mouse involves the activation of interneurons. *Neuroscience* **107**, 85–97 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

S.C.A., T.B. and G.R. contributed equally to this work. We thank S. Bourier, B. Brachvogel and W. Wurst for feeder cells and technical support; K. Pfeffer for E14 embryonic stem cells; K. Rajewsky for Cre deleter mouse line; B. Lüscher for FRT-flanked PGK-neo cassette; M. Wiedemann, H. Dietrich, B. Wölfl, A. Daschner, F. Fezza and A. Rippl for technical assistance, mouse breeding and genotyping; A. Mederer for help with behavioural experiments; F. Holsboer for continuous support; E. Güll for secretarial work; and C. Behl for critically reading the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.L. (e-mail: lutz@mpipsykl.mpg.de).

A transcription factor response element for gene expression during circadian night

Hiroki R. Ueda[†], Wenbin Chen^{*}, Akihito Adachi[‡], Hisanori Wakamatsu^{*}, Satoko Hayashi^{*}, Tomohiro Takasugi^{*}, Mamoru Nagano[‡], Ken-ichi Nakahama[‡], Yutaka Suzuki[§], Sumio Sugano[§], Masamitsu Iino[†], Yasufumi Shigeyoshi[‡] & Seichi Hashimoto^{*}

^{*} Molecular Medicine Laboratories, Institute for Drug Discovery Research, Yamanoichi Pharmaceutical Company, Ltd, 21 Miyukigaoka, Tsukuba, Ibaraki 305-8585, Japan

[†] Department of Pharmacology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Bunkyo-ku, Tokyo 113-0033, Japan

[‡] Department of Anatomy and Neurobiology, Kinki University School of Medicine, 377-2 Ohno-Higashi, Osakasayama City, Osaka 589-8511, Japan

[§] Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Mammalian circadian clocks consist of complex integrated feedback loops^{1–10} that cannot be elucidated without comprehensive measurement of system dynamics and determination of network structures¹¹. To dissect such a complicated system, we took a systems-biological approach based on genomic, molecular and cell biological techniques. We profiled suprachiasmatic nuclei and liver genome-wide expression patterns under light/dark cycles and constant darkness. We determined transcription start sites of human orthologues for newly identified cycling genes and then performed bioinformatical searches for relationships between time-of-day specific expression and transcription factor response elements around transcription start sites. Here we demonstrate the role of the Rev-Erba/ROR response element in gene expression during circadian night, which is in phase with *Bmal1* and in antiphase to *Per2* oscillations. This role was verified using an *in vitro* validation system, in which cultured fibroblasts transiently transfected with clock-controlled reporter vectors exhibited robust circadian bioluminescence¹².

To perform comprehensive measurement of mammalian circadian gene expression, we profiled genome-wide expression patterns of central (suprachiasmatic nuclei, SCN) and peripheral (liver) clocks every four hours during light/dark cycles (LD) or constant darkness (DD) over two days. We extracted total RNA from 50 pooled SCNs and four pooled livers at each time point, prepared biotinylated complementary RNA and used an Affymetrix mouse high-density oligonucleotide probe array (GeneChip) to determine SCN and liver gene expression.

The data obtained were analysed through two statistical cosine filters, one for LD and the other for DD time courses (see

Supplementary Information), to identify a set of genes rhythmically expressed under both LD and DD. On the basis of two successive filtrations, we classified 101 genes in the SCN and 393 genes in the liver as 'significantly rhythmic under both LD and DD'. In addition, we found two genes in the liver, *Rgs16* and *Hsp70-1*, with high 24-h autocorrelation under LD and DD and high cross-correlation between LD and DD liver expression profiles (see Supplementary Information), but without high correlation with the cosine curves. In total, we identified 101 genes in the SCN and 395 genes in the liver (Supplementary Information Tables 1 and 2). We also found 21 genes rhythmically expressed in both tissues (Supplementary Information Table 3). These are likely to be the minimum numbers of genes rhythmically expressed under both LD and DD in the SCN and/or liver; the number may increase if different filtration cut-off values are applied. Random profiles produced 18 and 13 samples classified as 'rhythmic under both LD and DD' in the SCN and liver, respectively. Using this estimate, we concluded that we had identified

a substantial number (~80 for the SCN and ~380 for the liver) of cycling genes.

Figures 1b and 2b display the cycling genes that we identified from the SCN and liver expression profiles, ordered by the averaged peak time between LD and DD (see Supplementary Information). Peak times of periodic expression profiles are spread over 24 h. When classified into 4-h interval groups ranging from CT0 (circadian time zero: the beginning of the subjective day) to CT24 (the end of subjective night, equal to CT0), we found the following distribution of peak times in the SCN (Supplementary Information Table 1): eight genes in CT0–4, 23 genes in CT4–8, 16 genes in CT8–12, 16 genes in CT12–16, 28 genes in CT16–20 and ten genes in CT20–24. We also found the following distribution in the liver (Supplementary Information Table 2): 91 genes in CT0–4, 71 genes in CT4–8, 58 genes in CT8–12, 63 genes in CT12–16, 50 genes in CT16–20 and 54 genes in CT20–24. The locations of these groups are indicated in Figs 1b and 2b.

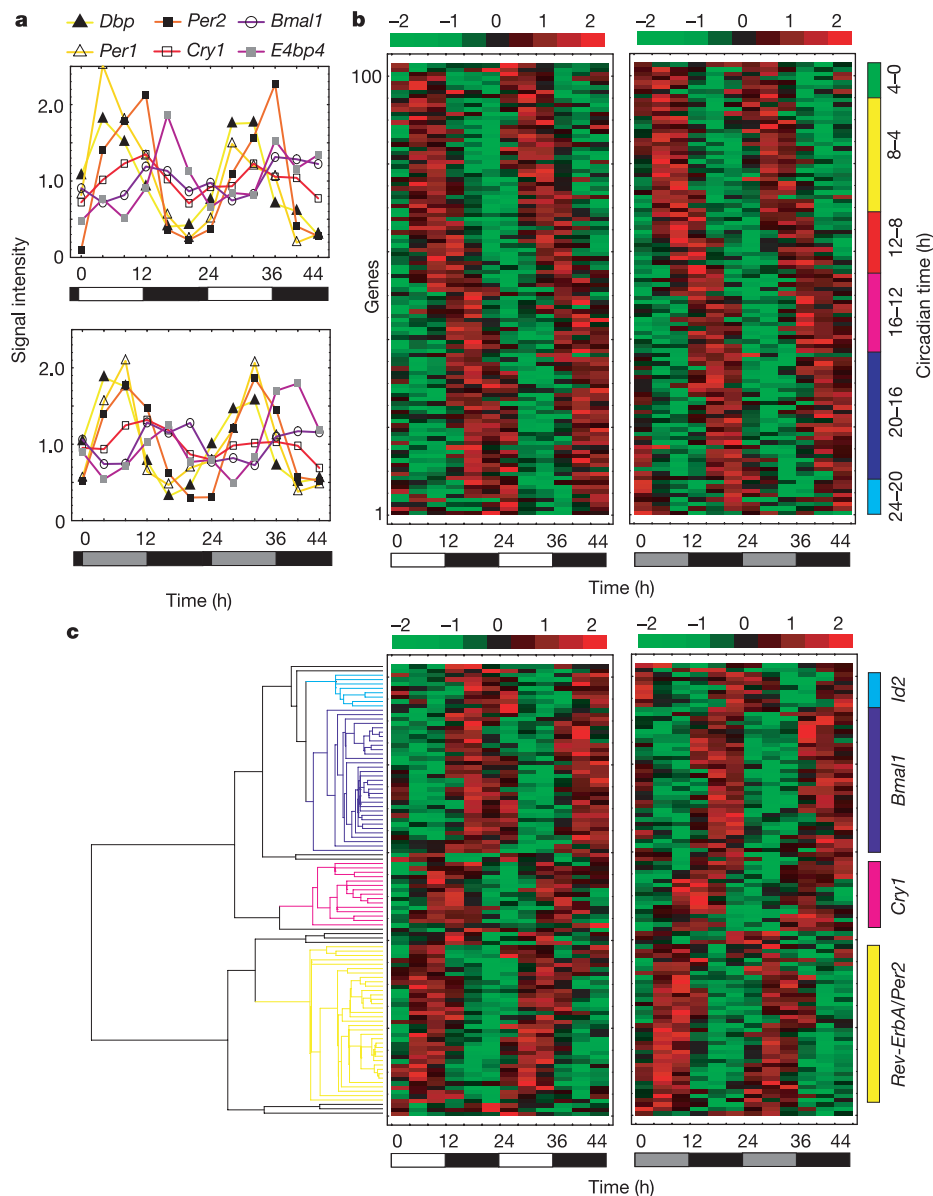


Figure 1 Temporal expression profiles in mouse suprachiasmatic nuclei. **a**, Periodic expression of known clock-controlled genes in mouse SCN under light/dark cycles (LD) and constant darkness (DD). Data were normalized so that the average signal intensity over 12-point time courses is 1.0. **b, c**, Organization of SCN expression profiles by peak

time (**b**) and hierarchical clustering (**c**). In both diagrams, columns represent time points, and rows represent genes. The colours in descending order from red to black to green represent the normalized data (the average and standard deviation over 12-point time courses are 0.0 and 1.0, respectively).

To cluster our cycling genes in accordance with the data, we performed hierarchical clustering. Figures 1c and 2c display the entire organization of our cycling genes in the SCN and liver, respectively. We found four main clusters in the SCN: the *Rev-ErbA/Per2* cluster (35 genes, average phase at CT6.0), *Cry1* cluster (15 genes, average phase at CT12.1), *Bmal1* cluster (32 genes, average phase at CT17.0) and *Id2* cluster (eight genes, average phase at CT20.8). We also found five main clusters in the liver: the *Rev-ErbA* cluster (87 genes, average phase at CT8.0), *Per2* cluster (75 genes, average phase at CT13.2), *Cry1* cluster (52 genes, average phase at CT17.8), *Bmal1* cluster (58 genes, average phase at CT20.5) and *Id* cluster (107 genes, average phase at CT3.1). The locations of these clusters are indicated in Figs 1c and 2c (see Supplementary Information Figs 1 and 2 for detailed results).

As an independent test, we measured the levels of expression under DD of three to 12 genes for each cluster using the quantitative polymerase chain reaction (PCR) assay in both SCN and liver. We

used the same samples that were used to prepare probes for microarray hybridizations. The expression profiles of the genes, as measured by these two independent methods, were very similar in both phase and amplitude (Supplementary Information Figs 3 and 4). These results confirmed our microarray measurements. To verify further our SCN microarray measurements, we explored the spatiotemporal expression profiles under DD of the genes identified in the SCN by an *in situ* hybridization study using ³⁵S-labelled cRNA probes (Fig. 3). The spatiotemporal expression profiles obtained from the *in situ* hybridization study were highly consistent with the temporal expression pattern obtained from the microarray experiments and further complement the microarray data because their signal intensity in the SCN is difficult to infer from microarray data.

Identification of more than 100 cycling genes in the SCN and nearly 400 cycling genes in the liver offered the opportunity to infer the transcriptional control mechanisms generating circadian expression. To do so, we analysed the promoter regions of newly

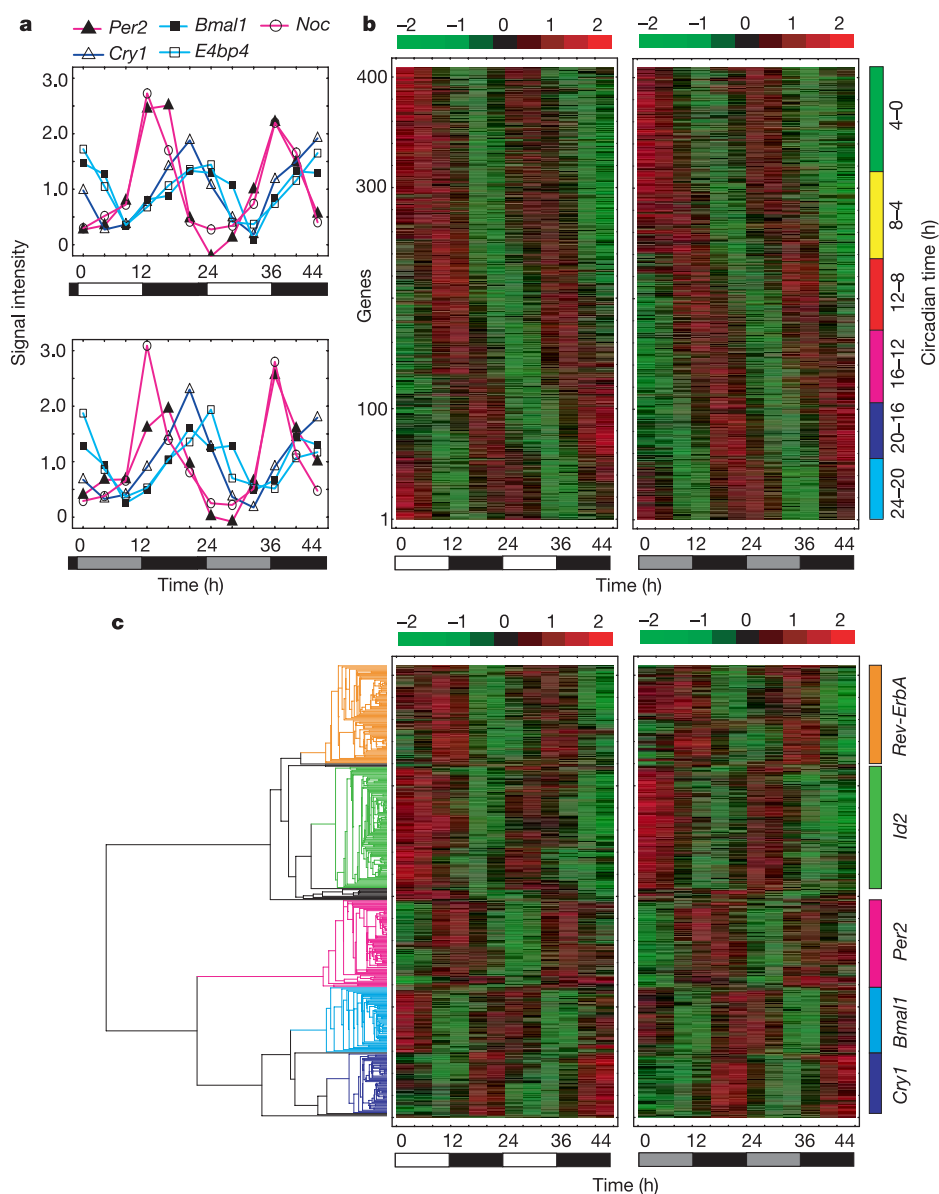


Figure 2 Temporal expression profiles in the mouse liver. **a**, Periodic expression of known clock-controlled genes in mouse liver under LD and DD. Data were normalized so that the average signal intensity over 12-point time courses is 1.0. **b**, **c**, Organization of liver expression profiles by peak time (**b**) and hierarchical clustering (**c**). In both diagrams,

columns represent time points, and rows represent genes. The colours in descending order from red to black to green represent the normalized data (the average and standard deviation over 12-point time courses are 0.0 and 1.0, respectively).

identified cycling genes in the SCN and liver. One of the main difficulties in analysing the promoter regions in higher eukaryotes, such as humans, mice and rats, is the lack of precise positional information of transcription start sites (TSSs). It is difficult to use translation start codons to estimate promoter regions because in many cases TSSs appear to be distant from translation initiation sites in higher eukaryotes. In fact, the TSS of the clock gene *Bmal1* is more than 60 kilobases (kb) away from the translation initiation ATG codon¹⁰. The other difficulty in analysing the promoter regions of mouse genes is the low coverage of publicly available sequence information over the entire mouse genome.

To address these two difficulties, we systematically determined the TSS of human orthologues of newly identified cycling genes by the oligo-capping method¹³ (see Methods) and then mapping the TSS onto the human genome, the draft sequence for which is publicly available. This approach is based on the evidence that circadian systems are well conserved between human and mouse¹⁴, and that the promoter region of the clock gene *Per1* is highly homologous in humans and mice¹⁵. We identified the TSS of 45 and 189 human orthologues for cycling genes in the SCN and liver, respectively (Supplementary Information Tables 4 and 5). Using positional information for these TSSs on the human genome, we defined the adjacent sequences as the putative promoter regions. We

then searched for the relationship between circadian expression and known circadian transcription factor response elements in the retrieved promoter regions of these genes (see Supplementary Information Tables 6 and 7 for detailed results).

We found seven cycling genes in the SCN with putative cyclic AMP response elements (CRE: TGACGT¹⁶) in the promoter regions of their orthologues, the phases of which consolidate to subjective day (circadian time, CT7.0 $\pm$ 1.6, mean $\pm$ s.d.). Phase consolidation of the genes with putative CRE is consistent with the previous observation that the reporter gene, driven by a CRE-containing promoter, displays clear circadian expression with a peak time occurring during subjective day in the SCN¹⁷.

We also found ten cycling genes in the SCN with putative Rev-ErbA/ROR response elements (AGGTCA¹⁸), to which Rev-ErbA and ROR family members bind, in the promoter regions of their orthologues. The phases of these genes consolidate to subjective night or dawn (CT19.2 $\pm$ 3.3, mean $\pm$ s.d.). The ten genes identified include *Bmal1* and *E4bp4*, which display similar circadian expressions antiphase to *Per2* oscillations in both SCN and the liver (Figs 1 and 2). These observations suggest that Rev-ErbA/ROR response elements play an important role in generating circadian oscillation antiphase to *Per* genes expression.

To explore the roles of Rev-ErbA/ROR response elements, we first examined in detail the spatiotemporal expression profiles in the SCN of all entire Rev-ErbA and ROR family members shown to bind to the same Rev-ErbA/ROR response element¹⁹. *Rev-ErbA*, *Rev-ErbA β* , *ROR α* and *ROR β* displayed similar circadian expression profiles in the SCN, with peaks during the day and troughs during the night (Fig. 4), whereas *ROR γ* was not detected in the SCN throughout the 24-h cycle (data not shown). *Rev-ErbA α* , *Rev-ErbA β* and *ROR α* demonstrated dynamic oscillation and were expressed very highly at CT4 (*Rev-ErbA α* and *Rev-ErbA β*) or CT8 (*ROR α*), with peak levels three times those of troughs. *ROR β* showed moderate to strong expression at every time point examined. The amplitude of *ROR β* oscillation was low compared with those of the other three genes in the family.

We also examined in detail the temporal expression profiles in the liver of the entire Rev-ErbA and ROR family by quantitative PCR. *Rev-ErbA α* , *Rev-ErbA β* and *ROR γ* exhibited clear circadian oscillations (Supplementary Information Fig. 5) whereas *ROR α* exhibited little or no circadian rhythmicity, and *ROR β* was not detected in the liver throughout the 24-h cycle (data not shown). *Rev-ErbA α* and *Rev-ErbA β* demonstrated dynamic oscillation, were very highly expressed at CT4–8 (*Rev-ErbA α*) and CT8 (*Rev-ErbA β*), and had peak levels more than 100 (*Rev-ErbA α*) and ten (*Rev-ErbA β*) times

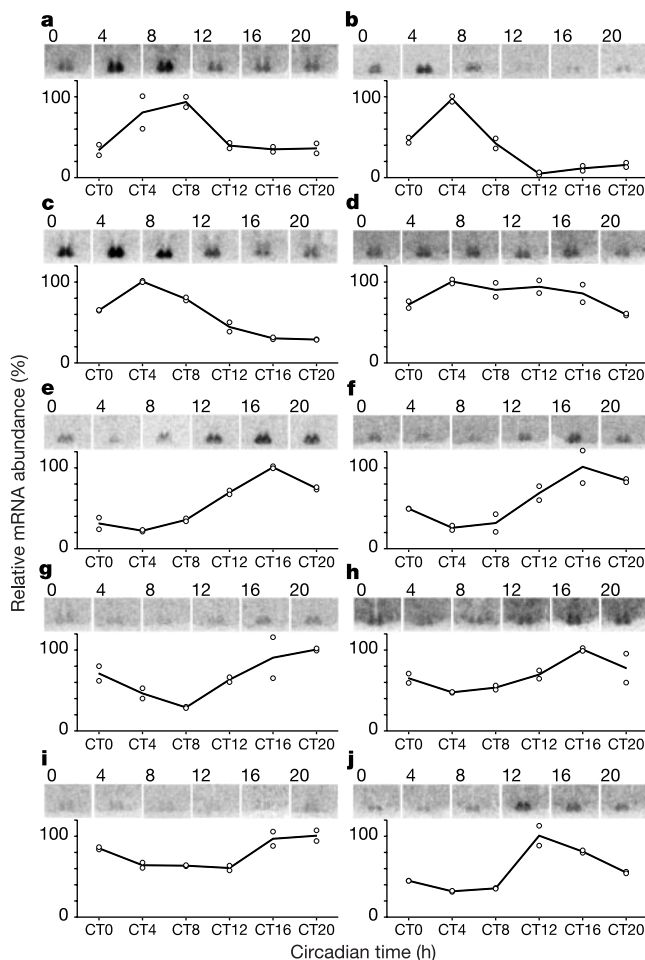


Figure 3 Spatiotemporal expression profiles in mouse SCN under DD conditions. Data for the genes (*Hmg4* (a), Ras-like protein 1190017B18Rik (b), *Rgs16* (c), Oligoribonuclease-like protein 1810038D15Rik (d), *Dextras1* (e), *V1a* (f), *Pki3* (g), *Hsp105* (h), *Id2* (i) and *Hsp40* (j)) consist of representative *in situ* hybridization autoradiograms of the SCN and quantitative graphs. Numbers above the radiograms indicate circadian time (h). Means are indicated by solid lines and individual values by open circles ($n = 2$).

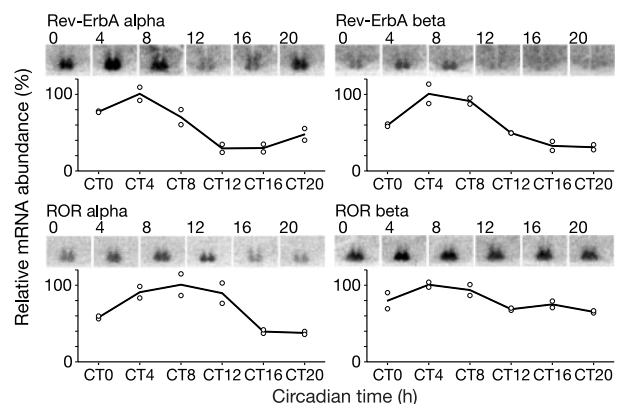


Figure 4 Spatiotemporal profiles of expression of the Rev-ErbA and ROR family in the mouse SCN under DD conditions. Representative *in situ* hybridization autoradiograms and graphs of relative mRNA abundance are shown. Numbers above the radiograms indicate circadian time (h). Means are indicated by solid lines and individual values by open circles ($n = 2$).

trough levels. *RORγ* also displayed clear oscillations, and was expressed highly at CT16–20 with a peak level three times the trough level.

To validate the hypothesized circadian circuit, we have developed an *in vitro* clock-controlled element determination system¹². We created *dLuc* by fusing a rapid degradation domain modified from mouse ornithine decarboxylase to the carboxy-terminal end of firefly luciferase. The rapid functional half-life of the *dLuc* proteins, 20–30 min, was suitable for probing circadian transcriptional dynamics in real time¹².

To verify the role of Rev-ErbA/ROR response elements in generating circadian expression antiphase to *Per2* oscillation, we fused the 60-base-pair promoter region of the *Bmal1* promoter, containing wild-type and mutant Rev-ErbA/ROR response elements, to the SV40 basic promoter driving a *dLuc* reporter gene (Fig. 5a). This 60-base-pair region is completely conserved between human and mouse *Bmal1* genes and is near the TSS in both cases. Transfection experiments confirmed that the wild-type construct can be activated by *RORα* and repressed by *Rev-ErbAα*, and that the mutant cannot (data not shown). To examine the circadian expression driven by the Rev-ErbA/ROR response element, we transfected these constructs into cultured Rat1-R12 fibroblasts, stimulated with dexamethasone, and measured luminescence from these cells. Bioluminescence levels from the wild-type Rev-ErbA/ROR response elements displayed circadian rhythmicity (Fig. 5d, *n* = 3) and the phase of their oscillations was in phase with the levels of luminescence from the *Bmal1* promoter fused to the destabilized luciferase (Fig. 5c, *n* = 3) and antiphase to those from the *Per2* promoter (Fig. 5b, *n* = 3). On the other hand, bioluminescence

levels from the mutant Rev-ErbA/ROR response elements exhibited no circadian rhythmicity (Fig. 5e, *n* = 3) and displayed light emission patterns similar to those from the SV40 promoter alone (Fig. 5f, *n* = 3). On the basis of these observations, we concluded that Rev-ErbA/ROR response elements play an important role in generating circadian night expression in phase with *Bmal1* expression and antiphase to *Per2* expression.

We have thus analysed mammalian circadian rhythms on the basis of genome-wide expression measurements, extensive determination of TSSs, bioinformatical search for transcription factor response elements, and *in vitro* validation of clock-controlled elements. The power of this systems-biological approach is demonstrated by the discovery and validation of the function of Rev-ErbA/ROR response elements in generating gene expression during circadian night, which is in phase with *Bmal1* and antiphase to *Per2* expression. This approach can be applied to the analysis of recently described mammalian cycling genes^{20–24}. Although efficient *in vitro* validation systems remain to be developed, the systems-biological approach may be applicable to circadian rhythms in higher eukaryotes such as *Drosophila*^{25–27} or *Arabidopsis*²⁸, as well as other biological systems such as mammalian cell cycles²⁹, in which genome-wide expression profiles have been measured. □

Methods

Microarray experiments

Slices (0.5 mm thick) of mouse brain were generated using Mouse Brain Matrix (Neuroscience) under LD or DD conditions, and then the SCNs were punched out bilaterally from the frozen slices under a stereomicroscope with a microdissecting needle (gauge 0.5 mm). Livers were dissected and frozen in liquid nitrogen. Total RNA was prepared from 50 pooled pairs of SCNs and four pooled livers at each time point using Trizol reagent (GIBCO BRL). Complementary DNA synthesis and cRNA labelling reactions were performed as previously described²⁶. We used Affymetrix high-density oligonucleotide arrays for *Mus musculus* (Murine Genome Array U74A) version 2 for SCN expression and version 1 for liver expression profiling. Expression data and other systems-biological information used in this paper will be available at the Database for Systems Biology (<http://www.dbsb.org/>).

Animals

Balb/c mice (male) purchased five weeks postpartum were adapted to standard 12-h light/12-h dark cycles (LD) for two weeks before samples were obtained under LD or constant darkness (DD) conditions every 4 h starting at ZT0 (zeitgeber time zero) or CT0 over two days.

In situ hybridization

The *in situ* hybridization and method of quantification of messenger RNA have been described in detail previously³⁰. Serial coronal sections (40 μm thick) of mouse brain were prepared using a cryostat. Fragments of cDNA were obtained by PCRs, which were then subcloned into PGEM-T easy vector (Promega). ³⁵S-labelled sense and antisense probes were transcribed using T7 or SP6 RNA polymerases. Antisense probes produced specific hybridization signals in the SCNs and/or other brain regions, whereas the sections hybridized with the sense probes exhibited no specific signal. The radioactivity of each section on BioMax film (Kodak) was analysed using a microcomputer interfaced to an image analysing system (MCID, Imaging Research) after conversion to relative optical density using ¹⁴C-acrylic standards (Amersham). Data were normalized with respect to the difference between signal intensities in equal areas of the SCN and the corpus callosum. The optical densities of the sections from the rostralmost caudalmost portion of the SCNs (ten sections per mouse) were then summed; the sum was used as the mRNA abundance in this region. 'Relative RNA abundance' means that the peak value was adjusted to 100%.

Quantitative PCR

To verify the microarray results, quantitative PCR was performed with the ABI Prism 7700 and SYBR Green Reagents (Applied Biosystems). cDNAs were synthesized from 0.5 μg of total RNA using Superscript II reverse transcriptase (Invitrogen). Samples contained 1 × SYBR Green Master Mix, 0.5 μM primers and 1/40 synthesized cDNA in a 25 μl volume. The PCR conditions were as follows: 10 min at 95 °C, then 40 cycles of 15 s at 94 °C, 30 s at 60 °C and 1 min at 72 °C. Absolute cDNA abundance was calculated using the standard curve obtained from mouse genomic DNAs. We used GAPDH as an internal control.

Determination of transcription start sites

In order to determine TSSs, oligo-capping was performed as described previously¹³ using 10 mg of polyA + RNA isolated from whole human brain, cerebellum and liver. cDNA was synthesized using a Thermoscript (Gibco BRL) from the oligo-capped mRNA and amplified by nested PCR with outer and inner 5' primers complementary to the cap-replacing oligo sequence and outer and inner 3' gene-specific primers. The PCR products were purified and sequenced. TSSs were identified precisely as the oligo-capped sites of the cDNAs.

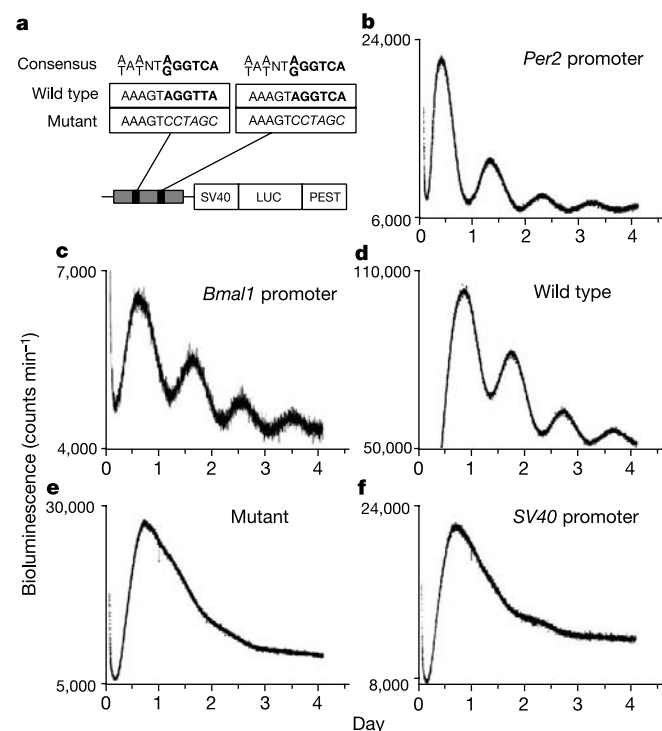


Figure 5 A Rev-ErbA/ROR response element sufficient for gene expression during circadian night. **a**, Diagram of wild-type and mutant 60-base-pair *Bmal1* promoter region fused to the SV40 basic promoter driving a destabilized luciferase (*dLuc*) reporter gene. The wild-type 60-base-pair region contains two putative Rev-ErbA/ROR response elements. Well-conserved base pairs¹⁸ (bold) and mutated base pairs (italic) are indicated. **b, c**, Representative circadian rhythms of bioluminescence from *Per2* (**b**) and *Bmal1* (**c**) promoter fused to a *dLuc* reporter gene. **d–f**, Representative circadian rhythms of bioluminescence from wild-type (**d**) and mutant (**e**) constructs or from SV40 promoter (**f**) fused to a *dLuc* reporter gene.

Constructs

The wild-type Rev-Erba/ROR response element and mutant Rev-Erba/ROR response element were constructed as follows: the following oligonucleotides were inserted into the *SmaI* site of the *SV40-dLuc* vector¹². WT: 5'-GATTGGTTCGGAAGTAGGTTAGTGGTGCGACATTAGGGAAGGCAGAAAGTAGGTCAGGGACGGAGG-3'. Mutant: 5'-GATTGGTTCGGAAGTCCTAGCGTGGTTCGACATTAGGGAAGGCAGAAAGTCCTAGCGGGACGGAGG-3'. The constructs were verified by sequencing.

Transfection and real-time monitoring of circadian bioluminescence

Rat1-R12 cells (ATCC) were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (Sigma). Cells were plated at 1.0×10^6 cells per dish in 35-mm dishes 24 h before transfection. Cells were transfected with LipofectAMINE 2000 reagent (GIBCO BRL) according to the manufacturer's instructions. Cells in each dish were transfected with 1 μ g (total) of expression plasmids. After 16 h, cells in each dish were treated with 0.1 μ M dexamethasone (Sigma), and after 2 h these media were replaced with 2 ml culture medium (Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (Sigma)) supplemented with 10 mM HEPES (pH 7.2), 0.1 mM luciferin (Promega), and antibiotics (25 U ml⁻¹ penicillin, 25 μ g ml⁻¹ streptomycin). Bioluminescence was measured with photomultiplier tube (PMT) detector assemblies (Hamamatsu). The modules and cultures were maintained in a light-tight incubator at 36 °C and interfaced to IBM PC-type computers for continuous data acquisition. The PMT was positioned about 2 cm above the culture, and photon counts were integrated over 1-min intervals.

Received 12 April; accepted 5 June 2002; doi:10.1038/nature00906.

- King, D. P. *et al.* Positional cloning of the mouse circadian clock gene. *Cell* **89**, 641–653 (1997).
- Bunger, M. K. *et al.* Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell* **103**, 1009–1017 (2000).
- Zheng, B. *et al.* Nonredundant roles of the mPer1 and mPer2 genes in the mammalian circadian clock. *Cell* **105**, 683–694 (2001).
- van der Horst, G. T. *et al.* Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* **398**, 627–630 (1999).
- Lowrey, P. L. *et al.* Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science* **288**, 483–492 (2000).
- Gekakis, N. *et al.* Role of the CLOCK protein in the mammalian circadian mechanism. *Science* **280**, 1564–1569 (1998).
- Jin, X. *et al.* A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. *Cell* **96**, 57–68 (1999).
- Kume, K. *et al.* mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* **98**, 193–205 (1999).
- Shearman, L. P. *et al.* Interacting molecular loops in the mammalian circadian clock. *Science* **288**, 1013–1019 (2000).
- Yu, W., Nomura, M. & Ikeda, M. Interactivating feedback loops within the mammalian clock: BMAL1 is negatively autoregulated and upregulated by CRY1, CRY2, and PER2. *Biochem. Biophys. Res. Commun.* **290**, 933–941 (2002).
- Kitano, H. Systems biology: a brief overview. *Science* **295**, 1662–1664 (2002).
- Chen, W., Ueda, H. R. & Hashimoto, S. In vitro determination of clock-controlled elements. *Curr. Biol.* (submitted).
- Maruyama, K. & Sugano, S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138**, 171–174 (1994).
- Clayton, J. D., Kyriacou, C. P. & Reppert, S. M. Keeping time with the human genome. *Nature* **409**, 829–831 (2001).
- Hida, A. *et al.* The human and mouse Period1 genes: five well-conserved E-boxes additively contribute to the enhancement of mPer1 transcription. *Genomics* **65**, 224–233 (2000).
- Paca-Uccaralarkun, S. *et al.* In vitro selection of DNA elements highly responsive to the human T-cell lymphotropic virus type I transcriptional activator. *Tax. Mol. Cell Biol.* **14**, 456–462 (1994).
- Obrietan, K., Impey, S., Smith, D., Athos, J. & Storm, D. R. Circadian regulation of cAMP response element-mediated gene expression in the suprachiasmatic nuclei. *J. Biol. Chem.* **274**, 17748–17756 (1999).
- Harding, H. P. & Lazar, M. A. The orphan receptor Rev-Erba alpha activates transcription via a novel response element. *Mol. Cell Biol.* **13**, 3113–3121 (1993).
- Jetten, A. M., Kurebayashi, S. & Ueda, E. The ROR nuclear orphan receptor subfamily: critical regulators of multiple biological processes. *Prog. Nucleic Acid Res. Mol. Biol.* **69**, 205–247 (2001).
- Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
- Duffield, G. E. *et al.* Circadian programs of transcriptional activation, signalling, and protein turnover revealed by microarray analysis of mammalian cells. *Curr. Biol.* **12**, 551–557 (2002).
- Akhtar, R. A. *et al.* Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr. Biol.* **12**, 540–550 (2002).
- Grundschober, C. *et al.* Circadian regulation of diverse gene products revealed by mRNA expression profiling of synchronized fibroblasts. *J. Biol. Chem.* **276**, 46751–46758 (2001).
- Storch, K. F. *et al.* Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78–83 (2002).
- McDonald, M. J. & Rosbash, M. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* **107**, 567–578 (2001).
- Ueda, H. R. *et al.* Genome-wide transcriptional orchestration of circadian rhythms in *Drosophila*. *J. Biol. Chem.* **277**, 14018–14052 (2002).
- Claridge-Chang, A. *et al.* Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron* **32**, 657–671 (2001).
- Harmer, S. L. *et al.* Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**, 2110–2113 (2000).
- Cho, R. J. *et al.* Transcriptional regulation and function during the human cell cycle. *Nature Genet.* **27**, 48–54 (2001).
- Shigeyoshi, Y. *et al.* Light-induced resetting of a mammalian circadian clock is associated with rapid induction of the mPer1 transcript. *Cell* **91**, 1043–1053 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank T. Kojima, T. Katakura and H. Urata for technical assistance. This study was performed as a part of a research and development project of the Industrial Science and Technology Program supported by NEDO (New Energy and Industrial Technology Development Organization).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.R.U. (e-mail: hiro@m.u-tokyo.ac.jp) or to Y.S. (e-mail: shigey@med.kindai.ac.jp).

A re-examination of proximodistal patterning during vertebrate limb development

Andrew T. Dudley*, María A. Ros† & Clifford J. Tabin*

* Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

† Departamento de Anatomía y Biología Celular, Facultad de Medicina, Universidad de Cantabria, 39011 Santander, Espana

The 'progress zone' model provides a framework for understanding progressive development of the vertebrate limb¹. This model holds that undifferentiated cells in a zone of fixed size at the distal tip of the limb bud (the progress zone) undergo a progressive change in positional information such that their specification is altered from more proximal to more distal fates. This positional change is thought to be driven by an internal clock that is kept active as long as the cells remain in the progress zone. However, owing to cell division, the most proximal of these cells are continually pushed outside the confines of the zone. As they exit, clock function ceases and cells become fixed with the positional value last attained while within the zone. In contrast to this model, our data suggest that the various limb segments are 'specified' early in limb development as distinct domains, with subsequent development involving expansion of these progenitor populations before differentiation. We also find, however, that the distal limb mesenchyme becomes progressively 'determined', that is, irreversibly fixed, to a progressively limited range of potential proximodistal fates.

Overlying the progress zone is a ridge of pseudostratified epithelium, known as the apical ectodermal ridge (AER), which runs along the anterior–posterior rim of the distal tip of the limb bud². When the AER is removed, the limb fails to form distal structures, suggesting that patterning of the distal limb is dependent on the AER³. The observation that less severe truncations result when the AER is removed at progressively later stages has been taken as strong evidence in support of the progress zone model. In this context, a 'specification map' for the formation of the skeletal elements of the chick wing has been proposed on the basis of examination of truncations resulting from AER removal at various stages⁴. We noted, however, that other reports have shown that, after removal of the AER, some distal mesenchyme cells are eliminated by cell death. In particular, one study detected a region of cell death extending 200 μ m from the distal tip within 4 h after AER removal at stage 22 (ref. 5). We reasoned that if this region of cell death remained constant in size at every stage, then a proportionally larger part of

the distal limb would be lost after removal of AER at early as opposed to late stages. Thus, AER removal could produce the observed pattern of distal truncation without reference to the progress zone or other models of proximodistal specification.

To test for this possibility, we collected wing buds 6–8 h after AER removal at various stages and assayed for apoptotic cells using a TdT-mediated dUTP nick end labelling (TUNEL) assay. Our results confirmed previous reports of cell death in a region extending up to 200 μm from the apex after AER removal at stage 22, and verified the previous report that cell death is no longer detectable 10 h after removal of AER⁵ (data not shown). Moreover, when we removed the AER between stages 18 and 22, the region of cell death remained similar in size (Fig. 1 and data not shown; $n = 3$ –5 per stage). At later stages, however, progressively less cell death was observed, with the region displaying significant cell death restricted distally by stage 24 (Fig. 1 and data not shown; $n = 3$). After stage 25, no cell death was detected (data not shown; $n = 3$).

Thus, although rapid cell death could potentially account for the progressive truncations seen after AER removal at early and mid stages of limb development, it cannot explain truncations within the

autopod (the most distal skeletal segment of the limb) when the AER is removed late in limb development. However, these late stage truncations could be explained if signals from the AER also affect proliferation of distal limb bud cells. To test this hypothesis, we assayed for cell proliferation by measuring 5-bromodeoxyuridine (BrdU) incorporation after AER removal. Indeed, we saw a marked loss of BrdU incorporation in the distal 200 μm of mesenchyme within 8 h after removing the AER from stage 24 wing buds (Fig. 1 and Table 1; $n = 3$).

In spite of these observations, the question remained whether the combination of cell death and decreased cell proliferation would be sufficient to explain the truncations seen after AER removal. We tested this by labelling distal cells and observing their fate after AER removal. At stage 19, only the stylopod (future humerus) forms after AER extirpation. According to the progress zone model, such truncations, which supposedly result from an arrest in the autonomous specification process thought to function within the progress zone, should allow for normal incorporation of labelled distal cells into the forming stylopod. On the other hand, if truncations arise because distal cells die, then distally labelled cells should mostly be absent after AER removal. Stage 19/20 ($n = 25$) and stage 23 ($n = 12$) limb buds were injected with Dil at 100 μm from the AER, and as a control, DiO was injected 300 μm below the AER. After a 24-h incubation, expansion of both proximal and distal cell populations was observed in control limbs containing an intact AER (Fig. 2a, c). In contrast, only a small number of labelled distal cells were found to survive after AER removal, and those that did remained tightly localized, whereas the proximal cell population expanded both in number and in distribution within the limb bud (Fig. 2b, d). Furthermore, when embryos that had undergone removal of the AER at stage 19 ($n = 12$) and stage 23 ($n = 9$) were allowed to incubate for 4 days, sections through these limbs demonstrated that proximal cells, but not distal cells, contributed to skeletal structures (Fig. 2e, f). These experiments demonstrate that distal cells do not contribute to skeletogenesis after AER removal. Thus, observing skeletal pattern after AER removal does not provide information regarding the specification of the distal cells at the time of surgery. Moreover, these data suggest that cell loss has a causal role in the resultant distal limb truncations.

Previous experiments have shown that application of fibroblast growth factor (FGF) protein can rescue proximodistal patterning when applied immediately after removal of the AER^{6,7}. Notably, FGF protein also prevents cell death in the distal limb bud mesenchyme after AER removal⁷. If these two results are causally related, then application of FGF protein after cells have died should fail to rescue truncations in AER-deleted early limb buds. Conversely, as a decrease in proliferation could be reversible, some distal pattern might be restored after a similar delay in FGF application when the AER is removed at later stages. To test this, we removed the AER and stapled heparin beads soaked in FGF4 to the distal tip of the operated limb, either immediately or after incubating the embryos an additional 12 h to allow for completion of cell death⁵. At stage 19/20, when appreciable levels of cell death are seen after AER removal,

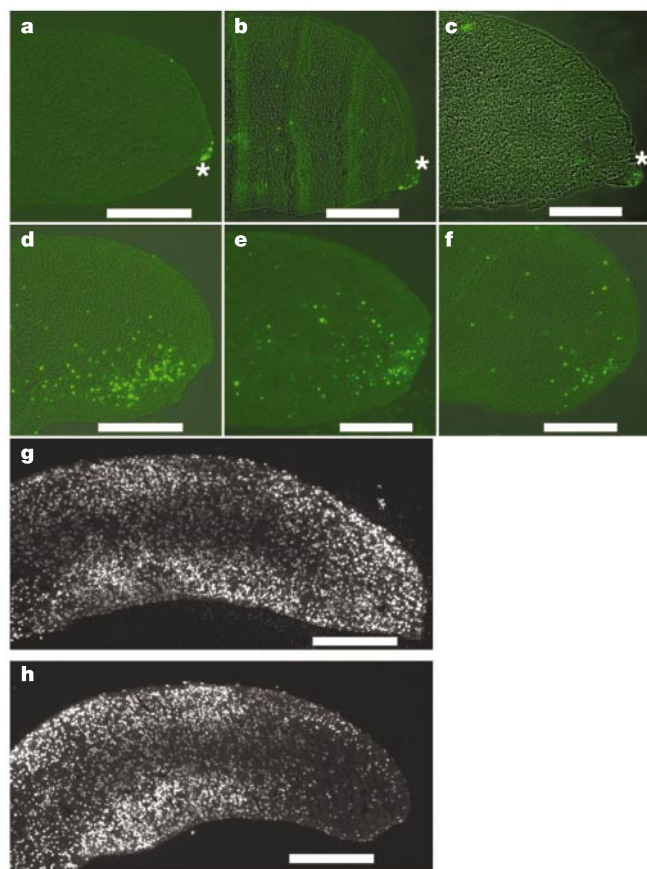


Figure 1 Removal of the AER results in the onset of cell death and a decrease in cell proliferation in the distal mesenchyme. **a–h**, Eight hours after AER removal, operated wing buds (**d–f, h**) and their contralateral controls (**a–c, g**) were sectioned and analysed for apoptotic cell death using the TUNEL reaction (**a–f**), or for cell proliferation by BrdU incorporation (**g, h**). Apoptotic cells are absent from the mesenchyme of intact limb buds (**a–c**), but are present in the AER (asterisk). Cell death extended at least 200 μm from the distal tip after AER removal at stage 19 (**d**) and stage 21 (**e**), and was still evident after AER removal at stage 24, but in a more restricted domain (**f**). At stage 24, BrdU positive cells are readily detected in the distal tip mesenchyme of unoperated limbs (**g**). After AER removal, significantly fewer distal cells display BrdU incorporation (**h**). Scale bars, 200 μm .

Table 1 Reduction in BrdU incorporation in the distal 200 μm of limb bud

Embryo	Treatment	n^*	BrdU positive nuclei (%)	Fold reduction†
1	Intact AER	6,033	35.4 (2.7)	
	Removed AER	9,911	17.0 (3.6)	2.1
2	Intact AER	8,220	40.3 (5.4)	
	Removed AER	8,930	14.3 (3.4)	2.8
3	Intact AER	7,097	41.0 (3.6)	
	Removed AER	9,069	13.2 (4.0)	3.1

Measurements were taken 8 h after removal of the AER at stage 24. Numbers in parentheses indicate standard deviation.

*Number of nuclei counted in each limb mesenchyme. Serial 6- μm paraffin sections were cut and every sixth section analysed. A minimum of 13 sections spanning the anterior–posterior axis were analysed for each limb.

†(Per cent positive nuclei for intact AER)/(per cent positive nuclei for removed AER).

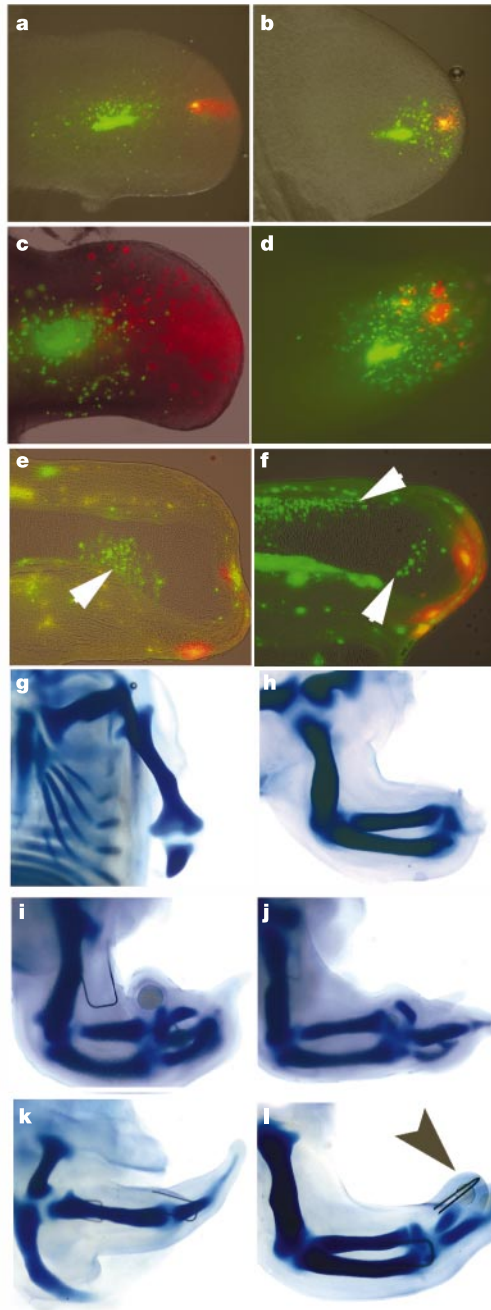


Figure 2 Cell death and decreased cell proliferation are significant factors in phenotypes produced by AER removal. **a–f**, Wing buds of stage 20 (**a, b, e**) or stage 23 (**c, d, f**) embryos were injected with both DiI (red) and DiO (green), 100 μ m and 300 μ m below the AER, respectively. After 20 h of incubation, both the proximal and distal cell populations expand in the unoperated limb buds (**a, c**); however, after AER removal, only the proximal cell population continues to expand significantly whereas the distal population fails to expand (**b, d**). Although some distal cells survive after AER removal (**e, f**), they do not incorporate into the limb skeleton after 4 days of development. Arrowheads mark incorporation of DiO into developing cartilage. **g–l**, Cell death is responsible for limb truncations after AER removal at early stages. After extirpation of the AER at stage 19 (**g, i, k**) or stage 23 (**h, j, l**), heparin beads soaked in FGF4 were stapled to the distal limb bud either immediately (**i, j**) or after 12 h (**k, l**). AER removal at stage 19 truncates development in the proximal part of the radius/ulna (**g**), whereas at stage 23 the limb skeleton develops up to the wrist (**h**). When FGF4 protein is added immediately after extirpation of the AER at either stage, a complete proximodistal set of skeletal elements develops (**i, j**). However, after the wave of cell death, FGF4 signalling can only rescue development in limbs of later stage embryos (compare **k** with **l**). The arrowhead in **l** denotes rudimentary digits. In each panel, distal is to the right. Magnification is 10 $\times$ (**a–f**) or 4 $\times$ (**g–l**).

FGF4 was able to rescue the proximodistal pattern only when added immediately after AER removal (Fig. 2g, i, k; $n = 14$). However, when the experiment was performed at stage 23, when less cell death is observed, FGF signalling was able to rescue proximodistal patterning defects even when applied 12 h after AER removal (Fig. 2h, j, l; $n = 8$).

Taken together, these data demonstrate that the domain of cell death is constant to a depth of approximately 200 μ m after removal of the AER at early stages of limb development; there is a substantial decrease in cell proliferation within the distal mesenchyme after AER removal at all limb bud stages analysed; and these effects result in loss of distal structures. Notably, examination of the most detailed fate maps available⁸ shows that removal of progenitor cells in the distal 200 μ m of the limb bud at various stages results in precisely the amount of distal deletion observed after extirpation of the AER at each of these stages. Therefore, we propose that the final skeletal pattern observed after AER removal does not reflect the level of patterning achieved through a particular stage of development. Rather, it reflects the proportion of progenitor cells which are outside the range of signals produced by the AER and are therefore not affected by the changes in cell death and proliferation caused by removal of the AER. Our data further emphasize the importance of the AER, and by extension FGFs, in preventing cell death and maintaining cell proliferation, as opposed to maintaining a cell autonomous patterning clock.

As the AER extirpation experiments do not give a reliable estimate of when proximodistal segments are specified during limb development, we turned to alternative approaches to address this issue. In a first set of experiments, we used the lipophilic dye DiI to examine the fates of cells in the early forelimb bud. Groups of neighbouring cells within a region less than 33 μ m in diameter were

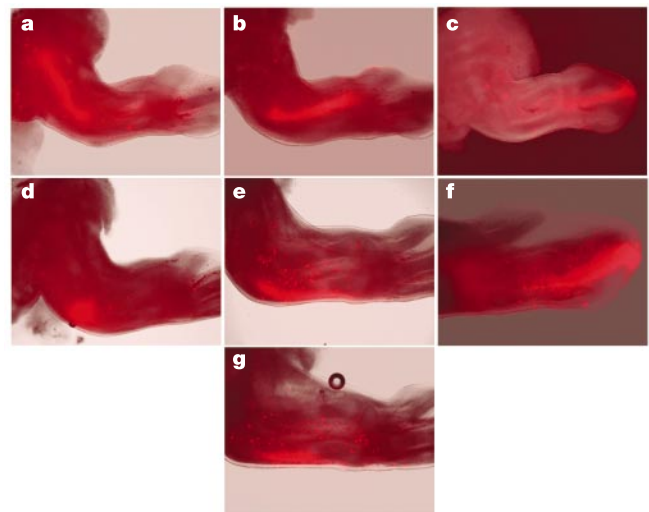


Figure 3 Descendants of cells in the early limb bud contribute to only one proximodistal segment. Wing buds of embryos were injected with the fluorescent lipophilic dye DiI and incubated for 3 days. **a–c**, Injections at the level of somite 18 directly below the AER labelled the autopod (**c**), injections 100–200 μ m below marked the zeugopod (**b**) and injections 200–300 μ m below marked the stylopod (**a**). **d–g**, Expansion of segment progenitors occurs in a proximal-to-distal sequence. Notably, DiI label applied at stage 19 extends the length of the respective segment, suggesting that expansion of the individual progenitor pools has not been completed by stage 19. However, DiI injections 200–300 μ m below the AER at stage 20 yield a discrete spot of label 3 days later, indicating that the stylopod progenitor population stops expanding by stage 20 (**d**), whereas the zeugopod progenitors are still expanding at stage 22 (**e**) but not by stage 23 (**g**). By contrast, injections subjacent to the AER at stage 26 result in label that extends throughout the phalanges, indicating that autopod progenitors have not finished expanding by stage 26 (**f**). Magnification is 4 $\times$.

labelled with DiI at stage 19 and assayed 3–4 days later. In most of the injections, labelled cells did not appear in more than one skeletal segment (Fig. 3a–c and Table 2; $n = 228$ out of 244). Thus, cells labelled directly under the AER marked the autopod (future digits), those labelled 100 μm beneath the AER marked the zeugopod (future radius/ulna), and those labelled 200 μm below the AER marked the stylopod. In a small number of cases, label was present in adjacent segments or was localized to the segmental boundary ($n = 16$ out of 244). We interpret these as representing cases where the injection site was at the border of two future segments and cells on both sides of the border were labelled. We also found that cells at the extreme anterior and posterior margins of the limb bud crossed the boundaries between proximodistal segments (data not shown), as has been noted previously⁹. However, as these cells are not part of the chondrogenic region¹⁰, it is not obvious that this finding is relevant to the discussion here.

Although DiI injections at different proximodistal levels at stage 19 were almost always confined to single limb segments, it was striking that these labelled cell populations consistently extended throughout the entire length of the segment (Fig. 3a–c). Thus, at stage 19, there is considerable mixing of cells along the proximodistal axis within each future segment, but not between segments. This suggests that the three limb segments are already specified and have different cellular properties at this early stage.

The spread of DiI in these limbs indicates that progenitor populations in all three segments undergo considerable expansion after injection. In contrast, at stage 20, we found that labelling cells with DiI at 250 μm below the AER resulted in a discrete spot of label confined to the stylopod but significantly smaller than the full limb segment (Fig. 3d; $n = 2$ out of 18 expanded). Thus, the stylopod progenitor cell population completes its expansion during the roughly 8 h between stage 19 and stage 20, consistent with the results of a previous report¹⁰ suggesting that proximal cell populations have fully expanded by stage 20. Similarly, we found that expansion of the zeugopod terminates between stages 22 (Fig. 3e; $n = 11$ out of 12 expanded) and 23 (Fig. 3g; $n = 2$ out of 12 expanded), whereas expansion of the autopod continues past stage 26 (Fig. 3f; $n = 6$ out of 9 expanded). In an accompanying manuscript, Sun *et al.* describe a model for AER-FGF function based on the concept that expansion of the stylopod, zeugopod and autopod progenitor populations is completed in such a proximal-to-distal sequence⁹.

To test directly whether the distal cells of the limb bud are specified to form autopod by stage 19, as the DiI labelling experiments would suggest, we transplanted the distal portion of stage 19 wing buds (fated to be autopod) to other sites in the embryo. When

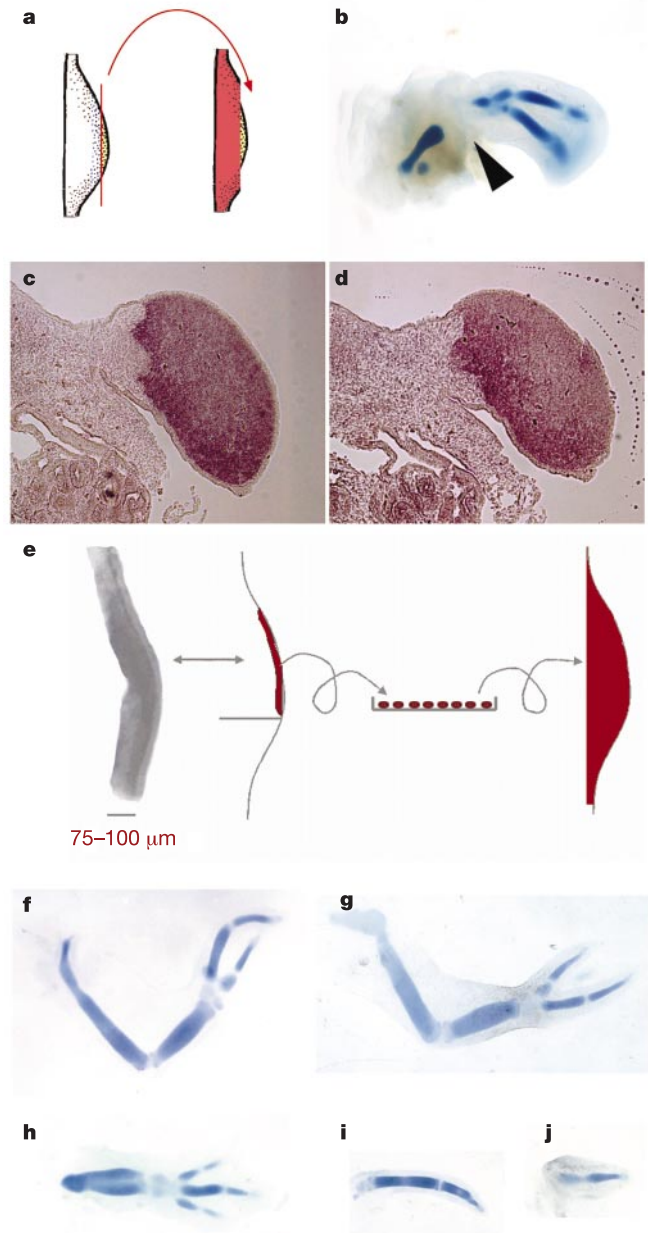


Figure 4 Distal cell fates are already specified but not determined in the early limb bud. **a**, Posteriorly biased slices from the distal tip of stage 19 wing buds approximately 100 μm thick were grafted to the stump of a donor leg bud. **b**, After 6 days of incubation, cartilage development was analysed by staining with alcian blue. Only digit-like elements formed from these grafts. The arrowhead denotes the graft–host interface. To verify that the transplanted cells maintained distal specification, some transplants were collected after 2–3 days, fixed and embedded in paraffin. **c**, **d**, *In situ* hybridization on 10- μm sections revealed that the population of transplanted cells, marked by wing-specific expression of *Tbx5* (**c**), correlates with expression of *Hoxa-13* (**d**), a marker of the future autopod. **e**, To test when distal cells are determined to be digits, a 75–100- μm slice of distal mesenchyme from the anterior two-thirds of the limb bud was disaggregated, re-aggregated, packed into an ectodermal hull and grafted to an embryo. After additional development, recombinant limbs were collected and the skeletal structure visualized using Victoria blue stain. **f**, **g**, Recombinant limbs generated from the entire anterior mesenchyme of stage 20 limbs (**f**) form all three proximodistal segments, as do limbs consisting of only distal anterior mesenchyme from stage 19 embryos (**g**). **h–j**, By stage 22, recombinants of distal mesenchyme form only the zeugopod and digits (**h**), whereas stage 23 recombinants are restricted to form digits (**i**), and stage 24 recombinants generate exclusively phalanges (**j**). Magnification is 4 $\times$ (**b**, **f–j**) or RELSP10 $\times$ (**c**, **d**).

Table 2 Proximodistal fate of cells DiI-labelled at various positions at stage 19

Position (μm) [*]	Somite no.†	n ‡	Wing-bud segment [§]					
			Autopod	A/Z	Zeugopod	Z/S	Stylopod	Other
U	17	25	22	3	–	–	–	–
	18	36	34	2	–	–	–	–
	19	31	29	2	–	–	–	–
100	17	20	–	–	18	2	–	–
	18	24	1	2	21	–	–	–
	19	19	17	2	–	–	–	–
200	17	12	–	–	–	–	12	–
	18	19	–	–	–	1	18	–
	19	21	–	1	18	1	1	–
300	17	11	–	–	–	–	1	10
	18	10	–	–	–	–	9	1
	19	16	–	–	–	–	14	2

^{*}Distance from the apex of the wing bud to the injection site. U, area directly underlying the AER.

†Somite opposite the injection site. Somite 17 is anterior to somite 19.

‡Number of limbs injected per injection site.

§The fate of labelled cells are indicated. Dye localized exclusively to one segment, except in cases where label was found at or spanning the border between the autopod and zeugopod (A/Z) or zeugopod and stylopod (Z/S).

||Label found adjacent to the limb bud but not obviously contained within the stylopod.

stage 19 distal wing tips were transplanted onto the stump of an amputated leg bud (Fig. 4a, b; $n = 9$) or into the coelom (data not shown; $n = 6$), the graft formed only digit-like skeletal elements. To verify the distal specification of the transplanted tissue, we examined molecular markers. Because a piece of the wing bud was grafted to the leg bud, tissue derived from the transplant could be unambiguously identified 2–3 days after transplantation using the wing-specific marker *Tbx5* (refs 11, 12) (Fig. 4c). In every case, the entirety of the graft expressed the autopod-specific marker *Hoxa-13* (ref. 13) (Fig. 4d; $n = 5$).

On the basis of our DiI labelling and transplantation experiments, we conclude that the proximodistal axis is patterned quite early in normal development. However, these studies do not address when cells of the various limb segments become determined, that is, refractory to re-specification to a different proximodistal fate under experimental conditions. Indeed, there is ample evidence indicating that early limb bud cells are not yet determined and therefore can change their proximodistal specification when challenged. For example, when limb mesenchyme is dissociated, re-aggregated, placed within an ectodermal hull, and then grafted back to a host embryo, the recombinant limb bud will develop and form skeletal structures^{14–16}. Although the posterior limb bud must be excluded because mixing polarizing cells with the recombinant mesoderm has a negative effect on its further development^{17,18}, if only the anterior two-thirds of the limb mesenchyme is used, all three limb segments can form relatively normal structures¹⁹ (Fig. 4f). Moreover, as cells from different proximodistal levels do not sort to their original position after re-aggregation¹⁵, the randomized organization of the recombinant limbs implies that cells formerly specified to form particular proximodistal structures contribute to different limb segments after this series of manipulations. We decided to use this model to examine when limb bud cells become determined with respect to their specific proximodistal fates.

When the entire stage-20 anterior limb bud mesenchyme was re-associated and allowed to develop, all three limb segments formed (Fig. 4f; $n = 12$ out of 12). We repeated this experiment using just the distal-most 75–100 μm of the anterior limb mesenchyme as a source of cells for the recombinant limb buds (Fig. 4e). A normal number of limb bud cells was achieved by pooling cells dissociated from the distal portion of multiple stage 20 limb buds. Importantly, these cells are already specified to form autopod structures according to the data described above. Moreover, according to the progress zone model, the distal limb bud cells have undergone significant proliferation between stages 17 and 20 and would be expected to have progressed beyond the specification of the stylopod. Nonetheless, recombinants made exclusively from the distal mesenchyme produce all three limb segments (Fig. 4g; 3 segments, $n = 6$ out of 9; 2 segments, $n = 3$ out of 9), showing that the distal cells are not yet determined and can be re-specified to form stylopod and zeugopod, as well as autopod.

At later stages, there is a progressive restriction in the ability of distal limb mesenchyme to produce proximal structures. Thus, at stage 22, recombinants formed zeugopod and digits but never a stylopod (Fig. 4h; 2 segments, $n = 1$ out of 2; 1 segment, $n = 1$ out of 2). Furthermore, at stage 23, recombinants were restricted in their competence to form autopod, including both metacarpals and phalanges (Fig. 4i; $n = 6$ out of 6). By stage 24, the distal mesenchyme only formed phalanges when re-aggregated (Fig. 4j; $n = 5$ out of 5). Similar results were seen at stage 26 (data not shown; $n = 4$ out of 4). Thus, there is a progressive restriction in the distal limb mesenchyme such that the cells are increasingly determined to form a narrower range of distal fates. This process of progressive determination takes place between stages 20 and 24.

In principle, the reaggregation data could also fit a progress zone model for specification if the distalization clock simply does not start ticking until stage 20. There is no reason to think this hypothetical clock would be inert during the earliest stages of

limb proliferation and then suddenly start after stage 20, but once one accepts the old AER-removal specification map as flawed (owing to cell death), then the timing of this process is totally unclear. It seems less *ad hoc* to consider specification of cellular properties reflected in proximodistal pattern, and the timing of determination, when those properties can no longer be altered, to be distinct and separable events. Proximodistal determination, or the stage when cells are no longer capable of being re-specified along this axis, is relevant not only to the re-aggregation experiments reported here, but also to a variety of other experimental manipulations that similarly lead to re-specification of limb mesenchyme, including grafting an ectopic AER proximally or transplanting a polarizing region anteriorly at different stages. All of these test the capacity of mesenchymal cells to be re-patterned along the proximodistal axis at various developmental stages, not the timing when that pattern is normally laid down.

Similarly, the demonstration by ref. 20 that X-ray irradiation of limb buds causes a more severe effect on proximal than distal limb development has been viewed as evidence for progressive specification of proximodistal pattern. However, we suggest that the implications of this experiment have been misunderstood because it has been interpreted in terms of the progress zone model of proximodistal specification when, more accurately, it addresses the degree to which the mesenchyme is determined at the time of irradiation. The progressive determination of distal fates is important as it limits the regulative ability of the undifferentiated tissue in the developing limb. In addition, differentiation of tissues within the limb proceeds in a proximodistal sequence such that, in the case of the wing, tissues of the shoulder and upper arm differentiate first³, whereas cells at the tip remain undifferentiated until later²¹. We suggest that these two properties, the progressive determination of undifferentiated cells and the proximal-to-distal wave of differentiation, together limit the ability of the developing limb to reconstitute pattern after cellular loss, and hence are probably responsible for the types of malformations seen after teratogenic events such as experimental irradiation and thalidomide exposure.

In contrast to the progressive nature of proximodistal determination, we find no evidence for progressive specification of fates as defined by the progress zone model. Rather, our results indicate that cell fates are established within the early limb bud at a time and over a spatial scale consistent with standard modes of inductive intercellular signalling. The realization that proximodistal specification and determination are distinct events will facilitate future work on the molecular and genetic bases for proximodistal limb patterning. □

Methods

Embryo manipulation

Fertilized chickens were obtained from Spafas and stored at 10 °C until use. To continue development, eggs were incubated at 38 °C in a non-rotating incubator. Eggs were windowed as previously described²². Embryos were staged according to ref. 23.

Before labelling cells with the fluorescent lipophilic dyes DiI and DiO (1,1-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate and 3,3'-dioctadecylcoxacarbocyanine perchlorate, catalogue number D-282 and D-275, respectively; Molecular Probes, 5 mg ml⁻¹ in dimethylformamide) the AER of the wing bud was lightly stained by pipetting a small amount of a 0.01% (w/v) solution of Nile blue sulphate in water against the distal tip of the limb bud. Dye was injected into wing buds *in ovo* using an Eppendorf pressure injector (Model 5242). Distance of the injection site from the AER was measured using a calibrated reticle on a Leica stereomicroscope. AER extirpation, if performed, occurred immediately after injection using tungsten needles. After incubation for 3–4 days, embryos were fixed in 1% paraformaldehyde/PBS, equilibrated in 30% sucrose/PBS, and photographed. Some embryos were subsequently embedded in OCT compound and sectioned on a cryostat. Digital images of fluorescent signal were taken on a Nikon Eclipse E1000 microscope using the Y-2 E/c Texas red (DiI) and fluorescein isothiocyanate/HYQ (DiO) filter sets.

For transplants, a 100- μm sliver of distal tip was cut from stage 19 wing buds opposite somites 17–19. This fragment was transplanted to a leg bud truncated approximately 100 μm distal to the flank. Amputation of the recipient leg bud close to the flank was necessary to avoid intercalary regeneration of zeugopod fates²⁴. The distal fragment was attached to the leg bud using 'U' shaped staples made from platinum wire (Goodfellow Metals; diameter 0.025 inches). After transplant, embryos were kept at room temperature for 1 h before returning to the incubator. In some cases the distal tips were grafted to the

coelom of stage 16/17 embryos by making a slit in the dorsal side of the embryo, between the prospective limb buds and lateral to the somites. The graft was inserted through the slit with the cut edge medial.

Recombinant limbs were prepared as described¹⁴, except that mesenchyme was derived from the distal-most 75–100 μm of the anterior two-thirds of the wing bud and that digestion with collagenase was omitted. Recombinant buds were grafted to the somites, allowed to develop for 7 days, and stained with Victoria blue as described¹⁹.

To rescue limb development in the absence of an AER, heparin acrylic beads were incubated in 1 mg ml^{-1} FGF4 (a gift from V. Rosen) at room temperature for 1 h, then stored on ice until use. After removal of the AER, two heparin beads were attached to the limb bud with platinum staples. One bead was placed at the posterior margin and the other just anterior to the first. After attachment of beads, embryos were kept at room temperature for 30 min before returning them to the incubator. Embryos were collected after 6 days of incubation and fixed in 4% paraformaldehyde. Skeletons were stained with 0.02% alcian blue 8GX in 70% ethanol/30% acetic acid at 37 °C, then cleared in 0.5% KOH and stored in glycerol.

Analysis of cell proliferation and death

After AER removal, embryos were returned to the incubator for 6–8 h. In cases where cell proliferation was examined, 200 μl of 5 mg ml^{-1} BrdU (Sigma) in PBS was injected around the embryo after 7 h, and the egg was returned to the incubator for 60 min. Embryos were collected, fixed in 4% paraformaldehyde/PBS overnight at 4 °C, washed in PBS, dehydrated through an ethanol series and embedded in paraffin. Sections (6 μm) were then floated onto TESPA (3-aminopropyltriethoxysilane)-treated slides, and apoptotic cells were stained using the method of ref. 25, except that biotin-16-dUTP was replaced by fluorescein-12-dUTP to allow for direct detection of transferred nucleotides. BrdU-labelled cells were detected using an anti-BrdU monoclonal antibody (clone BU-33, Sigma) followed by a Cy-2 conjugated goat anti-mouse secondary antibody (Jackson ImmunoResearch Laboratories). Slides for fluorescent microscopy were stained in 0.5 $\mu\text{g ml}^{-1}$ 4,6-diamidino-2-phenylindole (D-1306; Molecular Probes) in PBS for 1 min before mounting in 80% glycerol.

In situ hybridizations

Both the whole mount and non-radioactive section *in situ* hybridizations were performed using standard protocols for whole-mount *in situ* hybridization. Detailed protocols are available on request.

Received 19 March; accepted 2 June 2002; doi:10.1038/nature00945.

1. Summerbell, D., Lewis, J. H. & Wolpert, L. Positional information in chick limb morphogenesis. *Nature* **244**, 492–496 (1973).
2. Rubin, L. & Saunders, J. W. Ectodermal-mesodermal interactions in the growth of limb buds in the chick embryo: constancy and temporal limits of ectodermal induction. *Dev. Biol.* **28**, 94–112 (1972).
3. Saunders, J. W. The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *J. Exp. Zool.* **108**, 363–403 (1948).
4. Summerbell, D. A quantitative analysis of the effect of excision of the AER from the chick limb-bud. *J. Embryol. Exp. Morphol.* **32**, 651–660 (1974).
5. Rowe, D. A., Cairns, J. M. & Fallon, J. F. Spatial and temporal patterns of cell death in limb bud mesoderm after apical ectodermal ridge removal. *Dev. Biol.* **93**, 83–91 (1982).
6. Niswander, L., Tickle, C., Vogel, A., Booth, I. & Martin, G. R. FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell* (1993).
7. Fallon, J. F. *et al.* FGF-2: apical ectodermal ridge growth signal for chick limb development. *Science* **264**, 104–107 (1994).
8. Lewis, J. H. Fate maps and the pattern of cell division: a calculation for the chick wing bud. *J. Embryol. Exp. Morphol.* **33**, 419–434 (1975).
9. Sun, X., Mariani, F. V. & Martin, G. R. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* **418**, 501–509 (2002).
10. Vargesson, N., Clarke, J. D. W., Coles, V. K., Wolpert, L. & Tickle, C. Cell fate in the chick limb bud and relationship to gene expression. *Development* **124**, 1909–1918 (1997).
11. Gibson-Brown, J. J. *et al.* Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. *Mech. Dev.* **56**, 93–101 (1996).
12. Chapman, D. L. *et al.* Expression of the T-box family genes, Tbx1–Tbx5, during early mouse development. *Dev. Dyn.* **206**, 379–390 (1996).
13. Yokouchi, Y. *et al.* Misexpression of *Hoxa-13* induces cartilage homeotic transformation and changes cell adhesiveness in chick limb buds. *Genes Dev.* **9**, 2509–2522 (1995).
14. Ros, M. A., Lyons, G. E., Mackem, S. & Fallon, J. F. Recombinant limbs as a model to study homeobox gene regulation during limb development. *Dev. Biol.* **166**, 59–72 (1994).
15. Wada, N., Ohsugi, K., Yokouchi, Y., Kuroiwa, A. & Ide, H. Cell sorting and chondrogenesis aggregate formation in limb bud recombinants and in culture. *Dev. Growth Diff.* **35**, 421–430 (1998).
16. Hardy, A. *et al.* Gene expression, polarising activity and skeletal patterning in reagggregated hind limb mesenchyme. *Development* **121**, 4329–4337 (1995).
17. Crosby, G. M. & Fallon, J. F. Inhibitory effect on limb morphogenesis by cells of the polarizing zone coaggregated with pre- or postaxial wing bud mesoderm. *Dev. Biol.* **46**, 28–39 (1975).
18. Frederick, J. M. & Fallon, J. F. The proportion and distribution of polarizing zone cells causing morphogenetic inhibition when coaggregated with anterior half wing mesoderm in recombinant limbs. *J. Embryol. Exp. Morphol.* **67**, 13–25 (1982).
19. Piedra, M. E., Rivero, F. B., Fernandez-Teran, M. & Ros, M. A. Pattern formation and regulation of gene expressions in chick recombinant limbs. *Mech. Dev.* **90**, 167–179 (2000).
20. Wolpert, L., Tickle, C. & Sampford, M. The effect of cell killing by X-irradiation on pattern formation in the chick limb. *J. Embryol. Exp. Morphol.* **50**, 175–198 (1979).
21. Hornbruch, A. & Wolpert, L. Cell division in the early growth and morphogenesis of the chick limb. *Nature* **226**, 764–766 (1970).
22. Logan, M. & Tabin, C. Targeted gene misexpression in chick limb buds using avian replication-competent retroviruses. *Methods* **14**, 407–420 (1998).

23. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).
24. Kieny, M. & Pautou, M.-P. Proximo-distal pattern regulation in deficient avian limb buds. *Wilhelm Roux Arch.* **183**, 177–191 (1977).
25. Gavrieli, Y., Sherman, Y. & Ben-Sasson, S. A. Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* **119**, 493–501 (1992).

Acknowledgements

We thank J. Fallon for discussions about cell death; C. Tickle, G. Martin and J. Saunders for critiques of our data; A. Brent, N. Davis, C. Hartmann, G. Kardon, G. Martin, R. Pearce and K. Vogan for critical reading of this manuscript; and G. Martin and X. Sun for sharing data before publication. This work was supported by a grant from the National Institutes of Health (C.J.T.) and a fellowship from the Leukemia and Lymphoma Society (A.T.D.). M.A.R. was funded by the Spanish Ministry of Education and Culture.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.J.T. (e-mail: tabin@rascal.med.harvard.edu).

A first-generation linkage disequilibrium map of human chromosome 22

Elisabeth Dawson^{†‡}, Gonçalo R. Abecasis^{†‡§}, Suzannah Bumpstead^{*}, Yuan Chen^{*}, Sarah Hunt^{*}, David M. Beare^{*}, Jagjit Pabial^{*}, Thomas Dibling^{*}, Emma Tinsley^{*}, Susan Kirby^{*}, David Carter^{*}, Marianna Papaspyridonos^{*}, Simon Livingstone^{*}, Rocky Ganske^{||}, Elin Löhmusaar^{||}, Jana Zernant[#], Neeme Tõnisson[#], Maido Remm[¶], Reedik Mägi[#], Tarmo Puurand[¶], Jaak Vilo^{**}, Ants Kurg[¶], Kate Rice^{*}, Panos Deloukas^{*}, Richard Mott[‡], Andres Metspalu[¶], David R. Bentley^{*}, Lon R. Cardon[‡] & Ian Dunham^{*}

^{*} The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

[‡] Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

[§] Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, USA

^{||} Third Wave Technologies Inc., Madison, Wisconsin 53719-1256, USA

[¶] IMCB of the University of Tartu, and [☆] Estonian Biocentre, University of Tartu, 23 Riia Street, Tartu 51010, Estonia

[#] Asper Ltd., 3 Oru Street, Tartu 51014, Estonia

^{**} European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

[†] These authors contributed equally to this work

DNA sequence variants in specific genes or regions of the human genome are responsible for a variety of phenotypes such as disease risk or variable drug response¹. These variants can be investigated directly, or through their non-random associations with neighbouring markers (called linkage disequilibrium (LD))^{2–8}. Here we report measurement of LD along the complete sequence of human chromosome 22. Duplicate genotyping and analysis of 1,504 markers in Centre d'Etude du Polymorphisme Humain (CEPH) reference families at a median spacing of 15 kilobases (kb) reveals a highly variable pattern of LD along the chromosome, in which extensive regions of nearly complete LD up to 804 kb in length are interspersed with regions of little or no detectable LD. The LD patterns are replicated in a panel of unrelated UK Caucasians. There is a strong correlation between high LD and low recombination frequency in the extant genetic

map, suggesting that historical and contemporary recombination rates are similar. This study demonstrates the feasibility of developing genome-wide maps of LD.

Present-day chromosomes are mosaics of ancestral chromosomes that have arisen through multiple recombination events in the past. Each copy of a chromosome within a population can be uniquely characterized on the basis of a specific pattern of sequence variants, which together comprise an individual 'haplotype'. When these haplotypes do not occur in the population at the frequencies expected from the component variants, the variants are said to be associated or in linkage disequilibrium (LD). Characterizing the empirical patterns of LD across the genome will help to reconstruct the genetic history of human populations, enhance our understanding of the biological processes of recombination and natural selection, and facilitate association mapping studies that seek to localize genetic variants influencing complex traits and diseases. Previous studies of LD in humans have shown a high degree of variability, indicating that LD may extend between a few and several hundred kilobases^{1–5}. For practical applications, however, the local patterns of haplotype conservation are of primary interest^{6–8}, as the variability in LD overwhelms the average level. Therefore we characterized 59 independent haplotypes of human chromosome 22, derived from 77 members of three-generation pedigrees of the CEPH reference data set⁹. Using family samples helped to construct long haplotypes and detect genotyping error. All markers were selected from publicly available single-nucleotide polymorphisms (SNPs) and small insertions/deletions (indels)^{10–12} at regularly spaced intervals along the chromosome. A total of 951 of the 1,504 (63%) polymorphisms genotyped in the CEPH panels were common (minor allele frequency of ≥ 0.2 ; see Methods), which is similar to the proportion of common variants expected from comparing two chromosomes drawn from a constant size neutral

population (60%). LD between pairs of markers was calculated using the measure D' , following its usage in previous empirical studies, and the r^2 measure, which is preferred by population geneticists¹³.

LD decays with increasing distance, but also shows extensive variability (Fig. 1a, b). Maximal D' values extend up to distances over 400 kb, contrasting with occurrences of no detectable LD ($D' < 0.20$) between markers less than 5 kb apart. The distribution of r^2 values shows a similar degree of variability, differing from D' mainly in measurement scale (Spearman rank correlation $\rho(r^2, D') = 0.95$). Similar results were obtained in analysis of two separate populations of unrelated individuals, also of Caucasian origin, one from the UK (90 individuals) and one from Estonia (51 individuals; Fig. 1c, d). In the combined CEPH and UK data sets, average D' declines from 0.70 for adjacent markers to 0.11 for unlinked markers, whereas average r^2 declines from 0.38 to 0.01. Although the two measures differ in scale, their decay profiles are similar.

We assessed the pattern of LD along the chromosome by calculating average D' and r^2 for markers within contiguous 1.7-Mb stretches of DNA (sliding windows). In addition, statistical models of LD decay were fitted to summarize the patterns and to account for marker density. The results highlight areas with very high levels of LD, notably at positions 11–16 Mb and 21–27 Mb of the reference sequence (Fig. 2a, b). The degree to which these exceed background LD levels was confirmed by significance testing using extreme value distribution theory applied to 'runs' of LD (see Methods). Figure 2d shows various regions where LD is slightly above background levels (light bands), as well as shorter runs of extremely high disequilibrium relative to the rest of the chromosome (dark bands). These regional differences are not due to unequal marker spacing, as they align well with the model-fitting

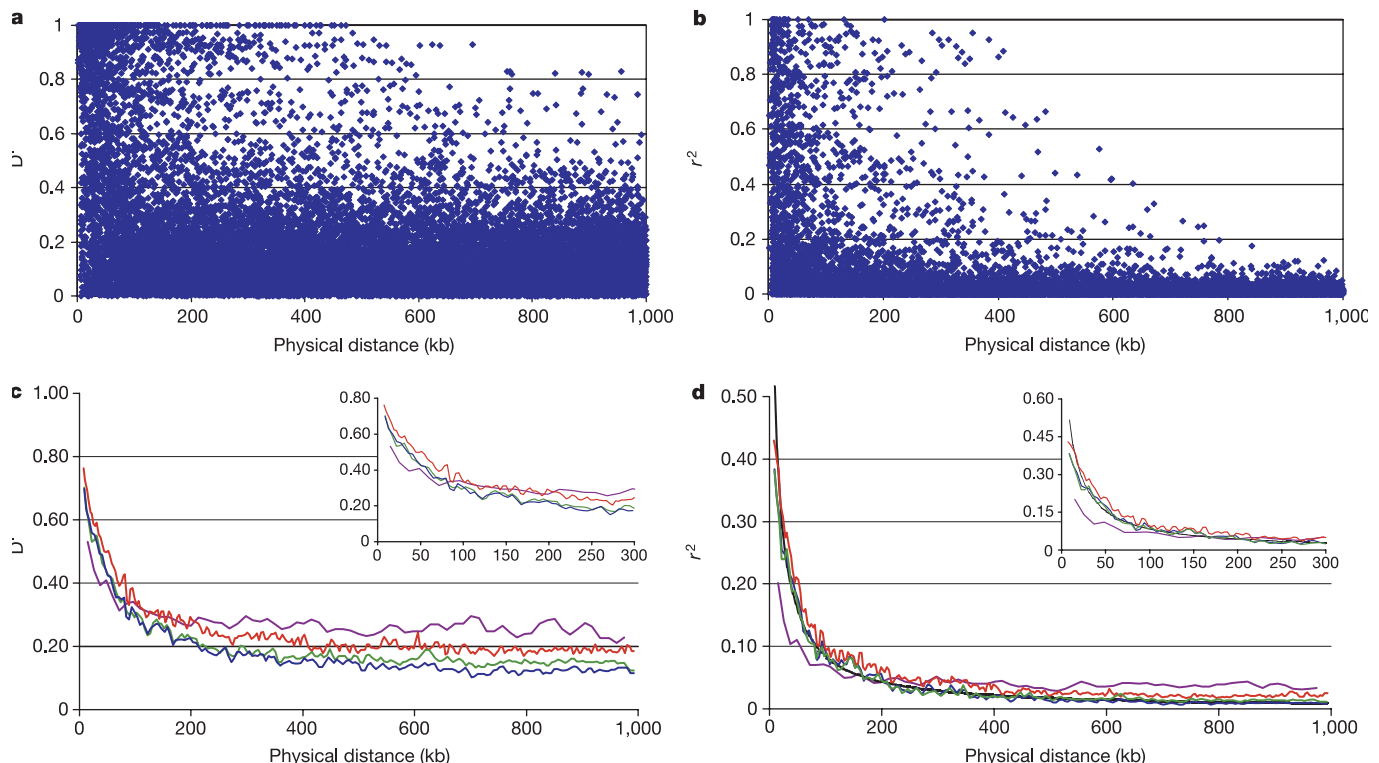


Figure 1 Distribution of linkage disequilibrium on chromosome 22. **a, b**, Variability of D' (**a**) and r^2 (**b**) for the CEPH samples, using all pairwise values for markers with minor allele frequency > 0.20 separated by ≤ 1 Mb. **c, d**, Sliding window results of average D' and r^2

in successive bins of 200 markers (100 marker overlap). Insets provide enhanced views of the observed LD decay from 0–300 kb. CEPH, red; unrelated UK, green; combined CEPH/UK, dark blue; Estonia, violet

results (Fig. 2b) that account for such variability. The estimates were consistent between the CEPH and UK samples (Fig. 2b), indicating the usefulness of both family-based and unrelated samples for initial detection of long LD runs. The general patterns of LD also appear similar in the Estonian samples (Fig. 1a), but the marker density in the smaller Estonian data set (median 34.72 kb) is too coarse for formal delineation of specific regions. In the CEPH and UK samples, average D' levels in the regions of high LD are 2–5 times greater than the background levels, presenting obvious distinctions of high and low LD tracts, whereas in the Estonian data, the highest LD region is less than twice the background level. Extrapolating these results to the genomic scale suggests that a median marker density greater than one marker per 35 kb is required for any first-generation map, and that, in the absence of family data, LD

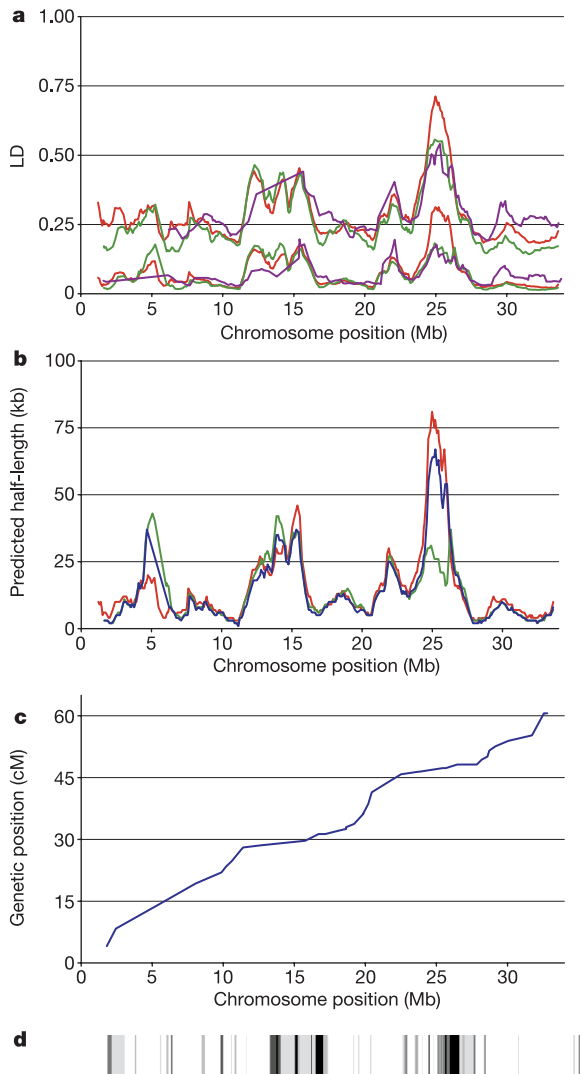


Figure 2 Linkage disequilibrium across chromosome 22. **a**, Average D' and r^2 coefficients (top and bottom groups, respectively) plotted in sliding windows containing all common polymorphisms separated by 50 and 500 kb in successive 1.7-Mb segments (1.6-Mb overlap). Sequence position 1 refers to the centromeric q-arm origin (ftp://ftp.sanger.ac.uk/pub/human/chr22/sequences/Chr_22/complete_sequence/Chr_22_19-05-2000.fa). The colour scheme is as in Fig. 1. **b**, Expected half length estimated from application of the model $E(r^2) = 1/(1 + 4Nc)$ to each sliding window. The model could not be fitted to the Estonian data because of the sparser marker density. **c**, Relationship between genetic and physical distance on chromosome 22 (refs 15, 16). **d**, Significant regions of excess LD (see Methods). The longest runs of LD (using $f = 0.50\sigma_D$) are shown in light grey; shorter runs of high LD are shown in dark grey ($f = 1.00\sigma_D$) and black ($f = 1.50\sigma_D$).

Table 1 Correlations between pairwise LD coefficients and sequence features				
Sequence feature	D'	r^2	$\hat{D}'_s$ (resid.)	$\hat{D}'_r$ (resid.)
Physical distance	-0.36	-0.37	0.05*	-0.11
Genetic distance	-0.50	-0.51	-0.30	0.02*
CpG coverage	0.04*	0.04*	0.15	0.00*
Coding coverage	0.12	0.12	0.19	0.05*
Gene coverage	0.13	0.13	0.20	0.06*
Pseudogene coverage	-0.10	-0.09	-0.02*	-0.06*
SINs	0.20	0.20	0.24	0.04*
LINEs	-0.12	-0.12	-0.09	0.00*
LTRs	-0.27	-0.26	-0.24	-0.15
DNA repeats	-0.13	-0.14	-0.07	-0.04*
Other repeats	-0.01*	-0.01*	0.11	0.05*

Correlations were Spearman rank correlations. CpG coverage, coding coverage, gene coverage and pseudogene coverage refer to the proportion of bases contained in predicted CpG islands, protein coding exons, all genic exons or pseudogenes, respectively, between each common SNP pair separated by ≤ 150 kb. All sequence characteristics were defined using the annotation in release 2.4 from the Sanger Centre chromosome 22 annotation group (http://www.sanger.ac.uk/HGP/Chr22/cwa_archive/Release_2.4_19-05-2000.shtml). The residual (resid.) columns represent deviates of pairwise linkage disequilibrium (LD) values from their expected values using exponential decay models⁴ fitted with sequence-based (bp) separation, $\hat{D}'_s$ (resid.) or genetic distance, $\hat{D}'_r$ (resid.). All correlations are significant at $P < 0.001$ except where noted by an asterisk. *Not significant.

estimates in 100 chromosomes or fewer may be too variable for reproducible patterns at a broad scale.

There is considerable evidence that sites of recombination in humans are not randomly distributed, but are often localized into specific hotspots^{7,14}. Current, low-resolution genetic maps can be used to model local recombination rates and provide additional predictors of LD beyond physical distance. Chromosome 22 has an elevated degree of recombination^{15–17}, averaging 2.46 cM Mb^{-1} in our interpolated sex-averaged genetic map (see Supplementary Information), compared with the genome average of approximately 1.3 cM Mb^{-1} . All components of the high LD tracts at 11–16 Mb and 21–27 Mb are situated in regions of exceptionally low recombination ($<1 \text{ cM Mb}^{-1}$; Fig. 2c) relative to the chromosome average. Indeed, nearly all of the high LD runs on chromosome 22 are located in regions of low recombination (Fig. 2d), a pattern previously noted in a localized region of this chromosome¹⁸, but which differs from a previous assessment of chromosome 22 microsatellite markers¹⁹. Collectively, the most exceptionally high-LD/low-recombination tracts cover about 9% of the chromosome, and 40% of the total variation in D' along chromosome 22 can be explained by the (interpolated) genetic distance between markers. These results indicate that the extant genetic maps may be used in practice as guides to genomic regions having high or low LD, and are of immediate use in position-based association studies.

To search for other predictors of LD, we examined correlations between pairwise LD measures and various sequence features (Table 1). LD is positively correlated with gene density and short interspersed nuclear elements (SINs) such as Alu repeats (which are features of (G + C)-rich sequence), and negatively correlated with long interspersed nuclear elements (LINEs), long terminal repeats (LTRs) and other DNA repeats ((G + C)-poor sequence). There is

Figure 3 Haplotype networks on chromosome 22. The figure is oriented with the leftmost position as the centromeric q-arm origin. The top panel shows the complete set of CEPH founder haplotypes (common alleles in blue; ambiguously phased or ungenotyped regions in white). Marker locations are given in the comb diagram underneath. The middle panels show 1-Mb-wide diagonal sections of colour-coded pairwise disequilibrium matrices³⁰ for the CEPH families (top), UK unrelated (middle) and combined samples (bottom), respectively. Each sample is followed by a comb diagram for marker location and a pictorial representation of haplotype networks. Networks of three or four markers are in blue; longer networks are in red. The bottom panels show a detailed view of common haplotypes ($>5\%$) for the combined data set. Within each block, haplotypes are listed in descending frequency. Gene composition and integrated genetic and physical maps are indicated. Transcribed sequences are grouped according to orientation¹⁵.

also a negative correlation between the (G + C)-rich sequence features and genetic distance. When genetic distance is factored out, by assessing sequence correlations with residual LD measures that are independent of genetic distance by multiple regression or using model-based regression residuals, nearly all of the relationships are reduced or effectively eliminated. The single exception to this collinearity occurs with LTRs, which maintain a significant negative correlation with LD independently of genetic distance ($\rho = -0.15$). It is conceivable that LTRs predict recombination at a relatively fine resolution, whereas the currently available coarse genetic map predicts recombination at a broader resolution.

In addition to pairwise disequilibrium assessments, our data provide an opportunity to identify common conserved haplotypes along the chromosome. We conducted a systematic search for regions of limited haplotype diversity and strong disequilibrium in the combined CEPH and UK samples, in which we detected 97 'haplotype networks' (see Methods), each including three or more markers (Fig. 3), including 329 out of 787 (41.8%) common polymorphisms and covering approximately 9.1 Mb of the long arm of chromosome 22 (22q). Interestingly, the two most common haplotypes are complementary at all sites in 55 of these networks. In some regions the networks overlap, reflecting the difficulty in precisely locating ancestral crossover events and possibly other phenomena such as gene conversion.

All of the common variants in any one of these networks can be surveyed with minimal genotyping effort. For example, in the CEPH families, the longest haplotype network extends 804 kb (at 11.83 Mb, including 16 markers) and a single network in the 21–27 Mb region includes 25 markers and extends 758 kb. These 25 markers could define up to 32 million chromosomes and in the absence of LD every chromosome in our sample would be unique. Instead, five haplotypes account for 76% of the CEPH founder chromosomes. Genotyping three SNP markers is sufficient to distinguish these five common haplotypes and retain 94.7% of haplotype heterozygosity. Similar long conserved haplotypes are present in the UK and combined samples.

Our ability to derive common haplotypes was significantly enhanced by the use of family-based samples (the CEPH panel). This is an important conclusion for a haplotype map project, because the relative usefulness of families will depend on empirical haplotype patterns²⁰. Comparison of the haplotype frequency estimates from the unrelated UK compared with partially-phased CEPH chromosomes indicated that, for an average set of five consecutive markers, each CEPH founder contributed 1.91 times more information than an unrelated individual (Methods). This suggests that although genotyping extended pedigrees such as the CEPH families requires twice as much effort as that for unrelated individuals, the information for LD mapping is approximately equal. Given this similarity and the advantages of families for detection of genotyping errors, integration with meiotic maps and quality comparison across laboratories, the use of families for LD mapping is clearly preferable to unrelated individuals.

The primary motivation for construction of any LD map in the human genome is to facilitate identification and characterization of genetic variants for common complex diseases. The present data indicate that considerable information is available for fine mapping of disease loci even in first-generation maps such as our 15-kb chromosome-wide resolution. For example, linkages to schizophrenia and other psychiatric disorders have been reported around 22q12, near marker D22S278 (ref. 21) (at approximately 19.8 Mb); schizophrenia has also been associated with microdeletions in 22q11, near the velo-cardiofacial syndrome locus (~5.6 Mb). Neither of these regions shows high LD in our data, suggesting that fine mapping may require a high density of markers. Conversely, linkage to type 2 diabetes has been reported around marker D22S423 (ref. 22) (approximately 23.8 Mb), which is on the edge of some of the longest tracts of high LD on the chromosome. Initial

allelic association in this region may be facilitated by this extensive conservation. □

Methods

Selection of markers, DNA samples and genotyping

Markers for genotyping were selected by walking along chromosome 22 in 15-kb steps through all available SNPs and small indels and choosing the nearest variant that was suitable for a unique polymerase chain reaction-based genotyping assay. Markers were genotyped in duplicate on 77 CEPH family DNAs and 90 unrelated UK Caucasian DNAs using the Third Wave Technologies Invader assay²³, and on 51 unrelated Estonian DNAs using allele specific primer extension in microarray format²⁴.

A final set of 1,504 markers were polymorphic in the CEPH DNA panel and were not rejected because of mendelian segregation errors, Hardy–Weinberg equilibrium deviations or other quality issues (the CEPH SNP set). There are 27 gaps of greater than 100 kb in this set (maximal gap of 293 kb), yielding a mean spacing of 22.95 kb and a median spacing of 15.07 kb. A total of 1,262 markers from the final CEPH set and 23 additional markers were successfully genotyped on the UK sample of unrelated individuals, for 1,286 markers on the UK Caucasians (the UK SNP set, median spacing of 17.53 kb, mean = 26.86 kb). We refer to the overlapping collection of CEPH and UK SNPs as the 'combined' marker set. The final Estonian SNP set had 908 SNPs, 661 of which had minor allele frequencies ≥ 0.20 , and a median spacing of 34.72 kb (mean = 61.42 kb). The Estonian SNP set included 594 SNPs in common with the initial CEPH SNP set. The final data used in these analyses is available at <http://www.sanger.ac.uk/HGP/Chr22>.

Error checking

We tested all markers for Hardy–Weinberg equilibrium, ignoring family structure, and excluded from the analysis those where equilibrium was rejected at the 10^{-4} level. We verified familial relationships within the CEPH samples and checked that presumed unrelated UK individuals were truly unrelated using the GRR program²⁵. In addition, we excluded all genotypes that produced mendelian errors or unlikely recombination patterns ($P < 0.001$) in the CEPH sample²⁶. The duplicate genotyping resulted in a very low error rate, as indicated by mendelian segregation errors (480 out of 98,095 genotypes = 0.5%) and unlikely double recombinants (315 out of 98,095 genotypes = 0.3%) in the CEPH data set.

Haplotyping

For the CEPH pedigrees, we used MERLIN²⁶ to list all alternate sets of non-recombinant founder haplotypes including small sets of neighbouring markers. Haplotype frequencies in families were then estimated using an expectation–maximization (E–M) algorithm²⁷ (software available on request from G.R.A.). For unrelated individuals and the combined data set, we estimated haplotype frequencies using the E–M algorithm.

Pairwise disequilibrium and distance modelling

For pairwise comparisons we calculated D' and r^2 following standard procedures¹³. We also fitted decay models to all pairwise coefficients within successive 1-Mb sliding windows: $E(r_{ij}^2) = 1/(1 + 4Nc_{ij})$, where N is the effective population size and c_{ij} is the recombination fraction between markers i and j estimated from the physical distance between markers using the chromosome 22 average 1 Mb $\approx$ 2 cM. We refer to the half-length of disequilibrium as the distance at which $E(r_{ij}^2) = 0.5$. Estimates from this model are largely independent of the underlying marker density.

Regions of excess disequilibrium

To define boundaries for regions of unusual disequilibrium, we used a method based on the Smith–Waterman algorithm²⁸. For the i th ordered marker pair within 500 kb of each other, we define the score $S_i = D'_i - k$ and then identify and compare high scoring segments using the Smith–Waterman accumulation approach and related statistical theory²⁹. We used a penalty $k = D' \times f \sigma_{D'}$ (with scale $f = 0.5, 1.0, 1.5$) to detect increasingly extreme runs of LD along the chromosome, where $\sigma_{D'}$ refers to the standard deviation of all pairwise LD coefficients.

Haplotype networks

We defined regions of limited haplotype diversity as those in which five haplotypes accounted for 75% or more of all haplotypes observed in the population and in which disequilibrium between each marker and haplotypes of surrounding markers exceed 0.75. We searched for sets of markers (networks) that met these conditions using a single marker as a seed and adding as many neighbouring markers as possible. We used MERLIN²⁶ to identify all non-recombinant haplotypes in a growing network and the E–M algorithm to estimate haplotype frequencies. We did not require markers in a network to be consecutive, but instead allowed up to six intervening markers to be excluded.

Comparison of families and unrelated individuals

Samples of unrelated individuals include more independent chromosomes, but less phase ambiguity exists in families. To compare the two approaches, we used the combined data set to estimate allele frequencies for each set of five consecutive markers. We then calculated the log-likelihood of each CEPH founder and unrelated individual using equilibrium allele frequencies and the haplotype frequencies estimated by E–M. The change in log-likelihood for each individual provides an indication of the amount of information contributed.

More detailed descriptions of marker selection, genotyping and error-checking

protocols, genotyped marker characteristics and statistical procedures are provided as Supplementary Information.

Received 31 December 2001; accepted 7 May 2002; doi:10.1038/nature00864.

Published online 23 June 2002; corrected 1 August 2002 (details online).

- Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22**, 139–144 (1999).
- Taillon-Miller, P. *et al.* Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nature Genet.* **25**, 324–328 (2000).
- Eaves, L. A. *et al.* The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nature Genet.* **25**, 320–323 (2000).
- Abecasis, G. R. *et al.* Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.* **68**, 191–197 (2001).
- Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
- Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. High-resolution haplotype structure in the human genome. *Nature Genet.* **29**, 229–232 (2001).
- Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
- Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
- Weissenbach, J. *et al.* A second-generation linkage map of the human genome. *Nature* **359**, 794–801 (1992).
- Dawson, E. *et al.* A SNP resource for human chromosome 22: extracting dense clusters of SNPs from the genomic sequence. *Genome Res.* **11**, 170–178 (2001).
- Mullikin, J. C. *et al.* An SNP map of human chromosome 22. *Nature* **407**, 516–520 (2000).
- Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
- Weir, B. S. *Genetic Data Analysis II* (Sinauer Associates, Sunderland, Massachusetts, 1996).
- Petes, T. D. Meiotic recombination hot spots and cold spots. *Nature Rev. Genet.* **2**, 360–369 (2001).
- Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
- Yu, A. *et al.* Comparison of human genetic and sequence-based physical maps. *Nature* **409**, 951–953 (2001).
- Payseur, B. A. & Nachman, M. W. Microsatellite variation and recombination rate in the human genome. *Genetics* **156**, 1285–1298 (2000).
- Eisenbarth, I., Striebel, A. M., Moschag, E., Vogel, W. & Assum, G. Long-range sequence composition mirrors linkage disequilibrium pattern in a 1.13 Mb region of human chromosome 22. *Hum. Mol. Genet.* **10**, 2833–2839 (2001).
- Huttley, G. A., Smith, M. W., Carrington, M. & O'Brien, S. J. A scan for linkage disequilibrium across the human genome. *Genetics* **152**, 1711–1722 (1999).
- Thompson, E. A., Deeb, S., Walker, D. & Motulsky, A. G. The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. *Am. J. Hum. Genet.* **42**, 113–124 (1988).
- Pulver, A. E. *et al.* Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12-q13.1: Part 1. *Am. J. Med. Genet.* **54**, 36–43 (1994).
- Ghosh, S. *et al.* The Finland–United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. *Am. J. Hum. Genet.* **67**, 1174–1185 (2000).
- Mein, C. A. *et al.* Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res.* **10**, 330–343 (2000).
- Kurg, A. *et al.* Arrayed primer extension: solid-phase four-colour DNA resequencing and mutation detection technology. *Genet. Test* **4**, 1–7 (2000).
- Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. GRR: graphical representation of relationship errors. *Bioinformatics* **17**, 742–743 (2001).
- Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30**, 97–101 (2001).
- Weir, B. S. & Cockerham, C. C. Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **42**, 105–111 (1979).
- Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981).
- Karlin, S. & Altschul, S. F. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl Acad. Sci. USA* **87**, 2264–2268 (1990).
- Abecasis, G. R. & Cookson, W. O. GOLD—graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).

Supplementary Information accompanies the paper on *Nature's* website <http://www.nature.com/nature>.

Acknowledgements

The authors thank the Wellcome Trust for support, A. Edwards for preparation of CEPH family DNA, J. Collins for assistance with chromosome 22 annotation, and M. Holgate for the genotype data extraction programs. We also thank E. Beatty and N. Jarvis. L.R.C. thanks the Wellcome Trust and the NIH for support. A.M. was supported by the Estonian Ministry of Education and an EstSF grant.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to I.D. (e-mail: id1@sanger.ac.uk) or L.C. (e-mail: lon.cardon@well.ox.ac.uk).

Golgi biogenesis in *Toxoplasma gondii*

Laurence Pelletier*, Charlene A. Stern†, Marc Pypaert*, David Sheff*, Huân M. Ngô‡, Nitin Roper*, Cynthia Y. He*, Ke Hu\$, Derek Toomre*, Isabelle Coppens‡, David S. Roos\$, Keith A. Joiner‡ & Graham Warren*

* Department of Cell Biology, Ludwig Institute for Cancer Research, Yale University School of Medicine, New Haven, Connecticut 06520, USA

† Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School and The Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA

‡ Section of Infectious Diseases, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06520, USA

\$ Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Two models have been put forward to explain the growth of new Golgi during the cell cycle. The first suggests that a new Golgi grows out of the endoplasmic reticulum by *de novo* synthesis¹. The second suggests that a pre-existing Golgi is needed for the growth of a new one, that is, the Golgi is an autonomously replicating organelle². To resolve this issue, we have exploited the simplicity of the apicomplexan parasite *Toxoplasma gondii*³, which has only a single Golgi stack⁴. Here we show, by using video fluorescence microscopy and three-dimensional reconstructions of serial thin sections, that the Golgi grows by a process of lateral extension followed by medial fission. Further fission leads to the inheritance by each daughter of a pair of Golgi structures, which then coalesce to re-form a single Golgi. Our results indicate that new Golgi grow by autonomous duplication and raise the possibility that the Golgi is a paired structure that is analogous to centrioles⁵.

The doubling of cell mass before division is accompanied by the duplication of cellular organelles. In the case of chromosomes and centrosomes, duplication is manifested as a doubling in number^{6,7}. In the case of membrane-bound organelles such as the endoplasmic reticulum (ER) and mitochondria, it is generally manifested as a doubling of mass because the number of these organelles varies, as does their unit size⁷. The Golgi apparatus may be an exception because it has a nearly constant unit size in all eukaryotes, comprising flattened cisternal membranes, typically 0.5–1.0 µm in cross-sectional diameter, that are most often arranged in a stack^{8,9}. Golgi duplication therefore results in a doubling of both number and mass.

Two different models have been proposed to explain the growth of new Golgi. The first suggests that new Golgi are assembled *de novo* from components that originate from the ER. The reversible absorption of the Golgi by the ER during mitosis or treatment with Brefeldin A (BFA) has been taken as evidence in favour of this view^{10,11}. The fact that cytoplasts lacking the Golgi cannot make a new one from the remaining ER is evidence against it¹². The second model suggests that new Golgi arise from pre-existing Golgi which either act as templates for assembly² or, by analogy with centrosomes⁵, instruct the assembly process. Morphological studies have provided images of Golgi, in organisms ranging from earthworms to plants, that are in the apparent act of medial fission (see ref. 7), but static images do not indicate the direction of events. Golgi in the act of separation could just as easily represent two Golgi in the act of fusing.

A significant problem in studying Golgi duplication is the number of Golgi in many cells. Mammalian cells typically contain 100–250 Golgi units, which are most often stitched together in a ribbon-like structure located near the cell nucleus^{13,14}. Even fungal cells, which are several orders of magnitude smaller in volume,

typically contain 5–20 discrete Golgi units per cell^{15,16}, which makes it difficult to track a single Golgi and observe the duplication process.

But many organisms, including protozoa, seem to have a single Golgi⁴. One of the best studied is *T. gondii*, an obligate intracellular parasite that invades host cells using a sophisticated array of secretory organelles and then replicates in a specialized parasitophorous vacuole¹⁷. The parasite's secretory organelles are supplied and maintained by a polarized secretory pathway comprising peripheral ER linked to the nuclear envelope, which contains an ER export site that supplies a single Golgi apparatus (ref. 18 and Fig. 1a). *Toxoplasma* uses the same core components and machinery used by mammalian cells for constitutive and regulated secretion, as well as other membrane traffic pathways³.

Electron microscopy images at different cell-cycle stages^{19,20} showed that the Golgi could be placed in one of four broad categories (Fig. 1a–d). Early in the cell cycle, there was a single Golgi with a cross-sectional diameter of about 0.7 μm (Fig. 1a). Later, the diameter had increased up to twofold, suggestive of lateral cisternal growth (Fig. 1b). Later still, there were two Golgi side-by-side, which suggested that medial fission had occurred (Fig. 1c). Last, the Golgi were incorporated into each nascent daughter cell formed by the process of endodyogeny, in which two daughter cells form within the mother (ref. 21 and Fig. 1d).

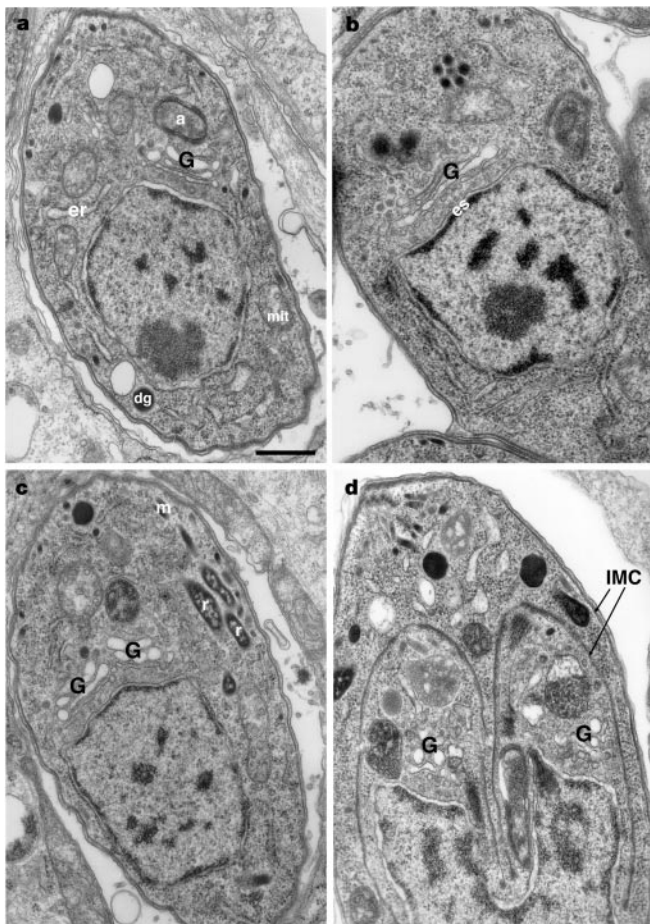


Figure 1 Stages of Golgi growth and division. Shown are thin section electron micrographs of *T. gondii* RH tachyzoites replicating by endodyogeny in HFF cells. Cells were placed in one of four categories according to the number and size of the Golgi: **a**, single Golgi; **b**, single, elongated Golgi; **c**, two Golgi; **d**, Golgi, often more vesiculated, in each nascent daughter cell, delineated by the growing inner membrane complex (IMC). See also the movies in Supplementary Information. a, apicoplast; dg, dense granules; er, ER; es, ER exit sites on the outer flattened part of the nuclear envelope; G, Golgi; m, micronemes; mit, mitochondria; r, rhoptries. Scale bar, 0.5 μm .

Because electron microscopy could not provide an unambiguous order of events, we confirmed this sequence by video fluorescence microscopy of live parasites using Golgi proteins tagged with spectral variants of green fluorescent protein (GFP). Mammalian GRASP55 was used as a representative structural or matrix protein²² because it has a close homologue in the *Toxoplasma* expressed sequence tag (EST) database. Stable expression of a fusion protein of yellow fluorescent protein (YFP) and mammalian GRASP55 (GRASP–YFP) showed discrete staining near the nucleus (Fig. 2a), in the same apical location as the Golgi identified in electron microscopy sections (Fig. 1). This location was confirmed by immunoelectron microscopy (Fig. 3a), and quantification showed that the density of labelling over the Golgi stack was at least five times higher than that in any other organelle (Fig. 3d, green). There was no detectable labelling over the cytoplasm, although low amounts were sometimes observed by fluorescence microscopy, especially in transiently overexpressing parasites, which is consistent with observations on tagged GRASP proteins in mammalian cells²³.

We used mammalian *N*-acetylglucosaminyltransferase I (NAGTI) as a representative Golgi enzyme marker¹³. There are no clear homologues of this or any other Golgi enzyme in the *Toxoplasma* EST database, which is consistent with the poor conservation of Golgi enzyme sequences across species²⁴, but there is strong biochemical evidence that complex oligosaccharides are constructed in *Toxoplasma*²⁵. NAGTI–YFP localized to the Golgi region in stably transfected *Toxoplasma* (Fig. 2b) and this location was confirmed by immunoelectron microscopy (Fig. 3b). Gold labelling in the Golgi was four times higher than that in the ER (Fig. 3d, red), and there was little if any labelling of other organelles including the parasitophorous vacuole. Occasional labelling of vacuoles by fluorescence microscopy (Fig. 2e, h, asterisks) suggested that some of this NAGTI construct was secreted, perhaps after being cleaved, but treatment with cycloheximide had little effect on Golgi labelling (Fig. 3b and data not shown), which showed that most of this construct was behaving as a resident Golgi protein.

Toxoplasma cells stably expressing both a cyan fluorescent protein (CFP) fusion of GRASP55 (GRASP–CFP) and NAGTI–YFP were also treated with BFA, which acts on protozoa as well as mammalian cells^{11,26}. NAGTI–YFP relocated from the Golgi to the nuclear envelope and a surrounding reticular structure (Fig. 2f–h). In

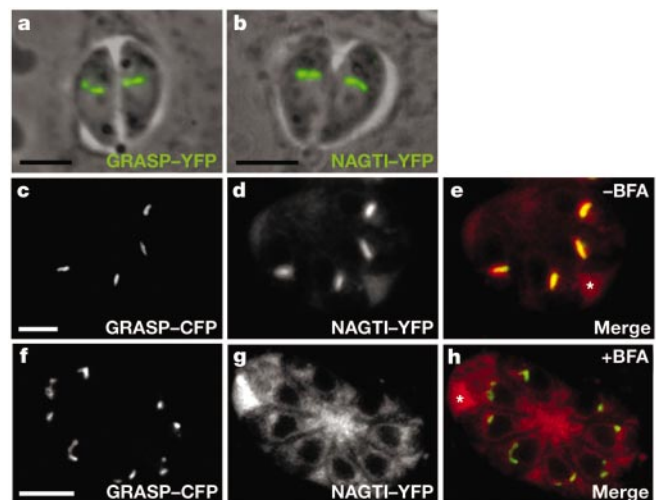


Figure 2 Stable expression of mammalian Golgi proteins. **a**, **b**, Overlaid immunofluorescence and phase images of GRASP–YFP (**a**) and NAGTI–YFP (**b**) in stable, transgenic cell lines of *Toxoplasma gondii*. **c**–**h**, Immunofluorescence images of a transgenic cell line expressing both GRASP–CFP (green) and NAGTI–YFP (red) before (**c**–**e**) or after (**f**–**h**) treatment with 5 $\mu\text{g ml}^{-1}$ BFA for 10 min at 37 $^{\circ}\text{C}$. Merged images are shown on the right. Asterisks indicate a secreted form of NAGTI–YFP that accumulates in the parasitophorous vacuole. Scale bars, 5 μm .

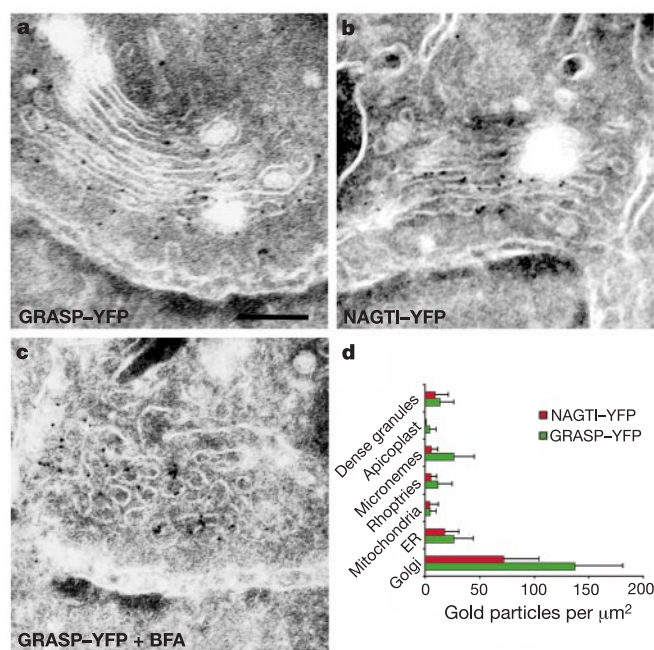


Figure 3 Immunoelectron microscopy of transgenic parasites. **a–c**, Cryosections of GRASP-YFP (**a**, **c**) or NAGTI-YFP (**b**) transgenic parasites, pretreated for 2 h with $50 \mu\text{g ml}^{-1}$ cycloheximide, before being fixed and immunolabelled for YFP using polyclonal antibodies against GFP followed by protein A coupled to 5-nm gold particles. Note the high density of labelling restricted to Golgi membranes. In **c**, GRASP-YFP transgenic parasites were treated with BFA ($5 \mu\text{g ml}^{-1}$) for 30 min before immunolabelling. Note the tubulo-vesicular appearance of the Golgi caused by loss of Golgi enzymes to the ER. **d**, Quantification of images in **a** and **b**. Results are presented as mean $\pm$ s.d. gold particles per μm^2 ($n = 20$ cells for GRASP-YFP; $n = 17$ cells for NAGTI-YFP). Scale bar, 200 nm.

separate experiments, this was shown to be the ER using cells expressing a *T. gondii* signal sequence, p30, fused to GFP and the ER retrieval signal, HDEL (ref. 18 and data not shown). In marked contrast, the location of GRASP-CFP was unaffected when assayed

using fluorescence microscopy (Fig. 2f–h), and immunoelectron microscopy of cells stably expressing GRASP-YFP (Fig. 3c) showed the tubuloreticular morphology of membranes left behind after loss of Golgi enzymes to the ER (see ref. 22). These data show that the Golgi in *Toxoplasma* can be fractionated by BFA and emphasize further the similarity between this organelle in *T. gondii* and mammalian cells.

Video sequences were filmed during a whole cell cycle, which lasts 6–8 h (ref. 20). To provide a marker for cell division, *Toxoplasma* was stably transfected with IMC1-CFP, a marker for the inner membrane complex (IMC)²¹. The IMC comprises a series of closed sacs that lie just beneath the mother cell plasma membrane and thus delineate the cell boundary (Fig. 1d). Each nascent daughter is demarcated by a growing IMC that packages the segregating cellular organelles (Fig. 1d). Eventually, daughter cells acquire their plasma membrane by budding out of the mother cell, which is consumed in the process.

These parasites were transiently transfected with GRASP-YFP and time-lapse images of infected cells were collected using epifluorescence microscopy. The video sequence is available as a QuickTime movie (Supplementary Information) and the key images are shown in Fig. 4. During G1, each cell contained what appeared to be a single Golgi structure (Fig. 4a). Over time, this structure grew (Fig. 4b) and then divided (Fig. 4c). Further division yielded four Golgi structures (Fig. 4d, e), with two partitioning into each daughter cell (Fig. 4e, f, arrows). Each daughter's pair of Golgi structures then appeared to coalesce and form a single structure (Fig. 4g, h).

Very similar results were obtained using the mammalian Golgi enzyme NAGTI-YFP, and double-labelling showed that this protein colocalized with GRASP-CFP at all stages of the cell cycle (Fig. 2c–e and data not shown). Notably, during cell division two Golgi structures containing NAGTI-YFP were inherited by each daughter (Fig. 4i, j), which showed that this pair contained both matrix and enzyme components.

The unexpected inheritance of a pair of Golgi structures prompted further ultrastructural examination of the Golgi in dividing cells. This was often more vesiculated (Fig. 1d), which made it difficult to define the boundary of the organelle in single

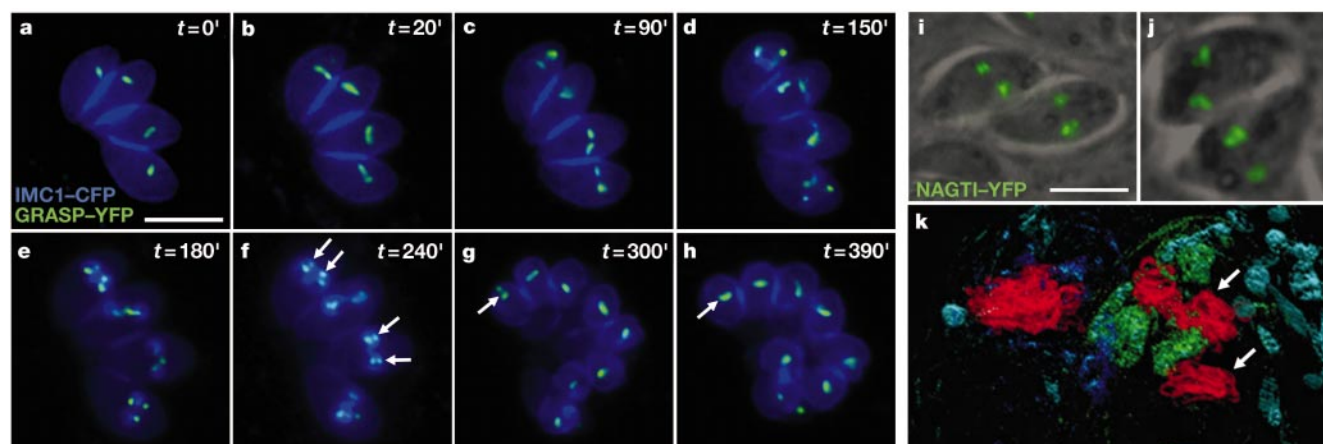


Figure 4 Biogenesis of the Golgi apparatus in living parasites. **a–h**, Transgenic parasites stably expressing IMC1-CFP (blue) were transfected with plasmid DNA encoding GRASP-YFP (green). After 20 h of infection in HFFs, four parasites were imaged by time-lapse video fluorescence microscopy. Images were taken every 10 min for 7 h at 37 °C. Representative images at the indicated times are shown. Note that *T. gondii* replicates synchronously in a given vacuole, which permits simultaneous imaging of several cells at the same cell-cycle stage. **i**, **j**, Transgenic parasites expressing NAGTI-YFP (green) were imaged over time and sample images (combined phase and fluorescence) late in cell division are shown. For both Golgi markers note the inheritance of two structures by each

nascent daughter (**f**, **i**, **j**) and their eventual coalescence (arrow in **g** and **h**). **k**, Three-dimensional reconstruction of two parasites during mitosis. The Golgi was selectively outlined in red and other electron-dense structures (such as dense granules) were coloured in green or dark blue to differentiate the two forming daughter cells. Golgi are inherited by both cells, and in the complete reconstruction of one daughter (right) two Golgi structures are visible (arrows). Note that the other daughter was only reconstructed partially and contains a single Golgi structure. See also the movies in Supplementary Information. Scale bar, 5 μm .

sections. We addressed this by reconstructing whole Golgi from serial thin sections and three-dimensional (3D) rendering, the latter being compiled into QuickTime movies (Supplementary Information). The representative view shown in Fig. 4k confirms the presence of two Golgi structures in a nascent daughter cell (the other daughter was only partially reconstructed and so contains only one). Other categories of Golgi shown in Fig. 1 were also reconstructed (Supplementary Information), both to confirm their overall size, shape and number, and to see whether reconstruction could reveal the presence of two substructures in a single Golgi. Despite several reconstructions, we have so far been unable to discern any consistent substructure, although there are hints from still and video fluorescence images that such substructures exist (Figs 2 and 4). The images did, however, clearly show the lateral growth and medial fission of the Golgi, as suggested by the single electron microscopy sections in Fig. 1.

Together, these experiments indicate that the Golgi in *T. gondii* may duplicate by lateral cisternal growth, followed by medial fission. All cisternae seem to grow as a cohort, and at the same rate as the ER export site on the nuclear envelope (Fig. 1a, b). There was no evidence that the ER export site grew first and then seeded the growth of Golgi *de novo*. This suggests that the growth of the Golgi and the export site are coordinated and that the two might even constitute a single integrated unit for duplication. Several workers have emphasized the intimate relationship between the Golgi and ER export sites in other organisms^{16,23}; therefore, *Toxoplasma* should provide an excellent model system in which to study this relationship and to try and understand the mechanism of duplication.

The inheritance of two Golgi structures by each daughter is both surprising and intriguing. Other protozoa, including *Trypanosoma brucei*, also seem to have more than two Golgi at certain life cycle stages²⁷, and it will be important to test the generality of this observation once these methods can be applied to more complex organisms such as mammalian cells. Inheritance of two Golgi structures might simply reflect the presence of two Golgi in each cell rather than one, but a more interesting explanation draws further on the analogy with centrosomes and the centriole pair that they contain⁵. In mammalian cells, the centriole pair duplicates in S phase and each pair is inherited by one daughter cell. Each member of the pair can, however, act independently and move to different parts of the cell at certain cell-cycle stages²⁸. Perhaps the Golgi is also a paired structure, which manifests itself most visibly in cell division. Perhaps, like centrioles, there are two very different functions that need to be carried out by each member of the pair that necessitates them being present as a paired organelle. Alternatively (or in addition), pairing may be inextricably tied up with the duplication process. It may be that autonomously replicating organelles have to be paired structures. □

Methods

Further details are available as Supplementary Information.

Cell culture and transgene expression

The general methods have been described²⁹. Human foreskin fibroblasts (HFFs) were infected with the *T. gondii* RH strain, which was transiently or stably transfected with CFP or YFP chimaeras of mammalian NAGTI and GRASP55 proteins.

DNA constructs

We used the *T. gondii* expression plasmids ptubFNR-YFP and ptubACP-CFP to derive plasmids encoding the amino-terminal 340 amino acids of rat GRASP55 and the N-terminal 103 amino acids (the retention signal) of human NAGTI linked to YFP or CFP. All constructs were verified by DNA sequencing.

Fluorescence microscopy and image analysis

Cells were imaged at 37 °C in a CO₂ perfusion chamber 18–24 h after infection using a Zeiss Axiovert 100M inverted microscope equipped with an Orca-100 CCD camera (1,280 × 1,022 pixels, 1 × 1 binning) and a Plan-Apochromat Ph3 100 × 1.4-NA objective. For simultaneous video imaging of CFP and YFP, we used an Olympus IX-70 inverted microscope equipped with a TILL Imago QE cooled CCD camera (1,376 × 1,040 pixels, 1 × 1 binning) and a 60 × 1.45-NA Plan-Apochromat objective.

Electron microscopy and 3D reconstitutions

For conventional or serial thin section electron microscopy, we fixed and processed HFF cells that had been infected for 30 h with wild-type RH parasites. Serial sections were scanned from *T. gondii* cells at different cell-cycle stages, manipulated in Adobe Photoshop and compiled into a single layered document. The layers were aligned to maximize the overlay of structures between serial sections; the Golgi apparatus was identified using morphological criteria and outlined in red. We imported the processed sections into Velocity for 3D rendering and generated QuickTime movies.

For immunoelectron microscopy, HFF cells were infected with transgenic parasites stably expressing either GRASP-YFP or NAGTI-YFP and then treated with 50 µg ml⁻¹ cycloheximide for 2 h before fixation, either with or without BFA (5 µg ml⁻¹) for the last 30 min. The cells were fixed and processed for immunoelectron microscopy using an affinity-purified polyclonal antibody against GFP, followed by protein-A-gold, and quantified using standard stereological techniques.

Received 22 April; accepted 12 June 2002; doi:10.1038/nature00946.

- Lippincott-Schwartz, J. & Zaal, K. J. Cell cycle maintenance and biogenesis of the Golgi complex. *Histochem. Cell. Biol.* **114**, 93–103 (2000).
- Bracker, C. E., Morre, D. J. & Grove, S. N. Structure, differentiation, and multiplication of Golgi-apparatus in fungal hyphae. *Protoplasma* **194**, 250–274 (1996).
- Joiner, K. A. & Roos, D. S. Secretory traffic in the eukaryotic parasite *Toxoplasma gondii*: less is more. *J. Cell Biol.* **157**, 557–563 (2002).
- Roos, D. S. et al. Transport and trafficking: *Toxoplasma* as a model for *Plasmodium*. *Novartis Found. Symp.* **226**, 176–195 (1999).
- Marshall, W. Centrioles take center stage. *Curr. Biol.* **11**, R487–R496 (2001).
- Kelly, T. J. & Brown, G. W. Regulation of chromosome replication. *Annu. Rev. Biochem.* **69**, 829–880 (2000).
- Warren, G. Membrane partitioning during cell division. *Annu. Rev. Biochem.* **62**, 323–348 (1993).
- Whaley, W. G. *The Golgi Apparatus* (Springer, New York, 1975).
- Preuss, D., Mulholland, J., Franzusoff, A., Segev, N. & Botstein, D. Characterization of the *Saccharomyces* Golgi complex through the cell cycle by immunoelectron microscopy. *Mol. Biol. Cell* **3**, 789–803 (1992).
- Zaal, K. J. et al. Golgi membranes are absorbed into and reemerge from the ER during mitosis. *Cell* **99**, 589–601 (1999).
- Presley, J. F. et al. Golgi membrane dynamics. *Mol. Biol. Cell* **9**, 1617–1626 (1998).
- Pelletier, L., Jokitalo, E. & Warren, G. The effect of Golgi depletion on exocytic transport. *Nature Cell Biol.* **2**, 840–846 (2000).
- Shima, D. T., Haldar, K., Pepperkok, R., Watson, R. & Warren, G. Partitioning of the Golgi apparatus during mitosis in living HeLa cells. *J. Cell Biol.* **137**, 1211–1228 (1997).
- Cole, N. B., Sciaky, N., Marotta, A., Song, J. & Lippincott-Schwartz, J. Golgi dispersal during microtubule disruption—regeneration of Golgi stacks at peripheral endoplasmic-reticulum exit sites. *Mol. Biol. Cell* **7**, 631–650 (1996).
- Wooding, S. & Pelham, H. R. The dynamics of Golgi protein traffic visualized in living yeast cells. *Mol. Biol. Cell* **9**, 2667–2680 (1998).
- Rossanese, O. W. et al. Golgi structure correlates with transitional endoplasmic reticulum organization in *Pichia pastoris* and *Saccharomyces cerevisiae*. *J. Cell Biol.* **145**, 69–81 (1999).
- Dubremetz, J. F., Garcia-Reguet, N., Conseil, V. & Fourmaux, M. N. Apical organelles and host-cell invasion by *Apicomplexa*. *Int. J. Parasitol.* **28**, 1007–1013 (1998).
- Hager, K. M., Striepen, B., Tilney, L. G. & Roos, D. S. The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii*. *J. Cell Sci.* **112**, 2631–2638 (1999).
- Sheffield, H. G. & Melton, M. L. The fine structure and reproduction of *Toxoplasma gondii*. *J. Parasitol.* **54**, 209–226 (1968).
- Radke, J. R. et al. Defining the cell cycle for the tachyzoite stage of *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **115**, 165–175 (2001).
- Hu, K. et al. Daughter cell assembly in the protozoan parasite *Toxoplasma gondii*. *Mol. Biol. Cell* **13**, 593–606 (2002).
- Seemann, J., Jokitalo, E., Pypaert, M. & Warren, G. Matrix proteins can generate the higher order architecture of the Golgi apparatus. *Nature* **407**, 1022–1026 (2000).
- Ward, T. H., Polishchuk, R. S., Caplan, S., Hirschberg, K. & Lippincott-Schwartz, J. Maintenance of Golgi structure and function depends on the integrity of ER export. *J. Cell Biol.* **155**, 557–570 (2001).
- Wiggins, C. A. R. & Munro, S. Activity of the yeast mnn1 α -1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc. Natl Acad. Sci. USA* **95**, 7945–7950 (1998).
- Odenthal-Schnittler, M., Tomavo, S., Becker, D., Dubremetz, J. F. & Schwarz, R. T. Evidence for N-linked glycosylation in *Toxoplasma gondii*. *Biochem. J.* **291**, 713–721 (1993).
- Lujan, H. D. et al. Developmental induction of Golgi structure and function in the primitive eukaryote *Giardia lamblia*. *J. Biol. Chem.* **270**, 4612–4618 (1995).
- Field, H., Sherwin, T., Smith, A. C., Gull, K. & Field, M. C. Cell-cycle and developmental regulation of TBRAB31 localisation, a GTP-locked Rab protein from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **106**, 21–35 (2000); erratum **107**, 329–330 (2000).
- Piel, M., Nordberg, J., Euteneuer, U. & Bornens, M. Centrosome-dependent exit of cytokinesis in animal cells. *Science* **291**, 1550–1553 (2001).
- Roos, D. S., Donald, R. G. K., Morrisette, N. S. & Moulton, A. L. C. Molecular tools for genetic dissection of the protozoan parasite *Toxoplasma gondii*. *Methods Cell Biol.* **45**, 27–63 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank F. Barr and T. Nilsson for the GRASP55 and NAGTI plasmids; S. Trombetta for recombinant GFP; the Joiner and Roos labs—mainly F. Quittat, T. Stedman and O. Harb—for advice and discussion; N. Andrews for the use of her cell-culture room;

C. Roy and J. Kagan for help with confocal microscopy; K. Zichichi, C. Horensavitz and C. Deloye-Pypaert for help with electron microscopy; Olympus for providing some of the microscopes; J. Shorter and E. Procyk for critically reading the manuscript; and the Warren and Mellman labs for discussions. This work was supported by National Institutes of Health grants (to G.W., I.M., K.A.J. and D.S.R.) and the Ludwig Institute for Cancer Research.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and request for materials should be addressed to G.W. (e-mail: graham.warren@yale.edu).

Two-step binding mechanism for T-cell receptor recognition of peptide–MHC

Lawren C. Wu^{*,†}, Delphine S. Tuot^{*}, Daniel S. Lyons^{*,†}, K. Christopher Garcia[†] & Mark M. Davis^{*,†}

^{*} Howard Hughes Medical Institute, [†] Department of Microbiology and Immunology, Stanford University School of Medicine, B221 Beckman Center, Stanford, California 94305, USA

T cells probe a diverse milieu of peptides presented by molecules of the major histocompatibility complex (MHC) by using the T-cell receptor (TCR) to scan these ligands with high sensitivity and specificity¹. Here we describe a physical basis for this scanning process by studying the residues involved in both the initial association and the stable binding of TCR to peptide–MHC, using the well-characterized TCR and peptide–MHC pair of 2B4 and MCC-IE^k (moth cytochrome *c*, residues 88–103)². We show that MHC contacts dictate the initial association, guiding TCR docking in a way that is mainly independent of the peptide. Subsequently, MCC-IE^k peptide contacts dominate stabilization, imparting specificity and influencing T-cell activation by modulating the duration of binding. This functional subdivision of the peptide–MHC ligand suggests that a two-step process for TCR recognition facilitates the efficient scanning of diverse peptide–MHC complexes on the surface of cells and also makes TCRs inherently crossreactive towards different peptides bound by the same MHC.

The $\alpha\beta$ TCR is a heterodimeric molecule consisting of variable and constant immunoglobulin domains in which the complementarity determining regions (CDRs) of the variable domains comprise the peptide–MHC binding surface³. The TCR binds peptide–MHC in a conserved diagonal orientation that positions the CDR1 and CDR2 loops mainly over the MHC and the CDR3 loops over peptide. TCR recognition is highly sensitive to small numbers of ligands among thousands of different peptide–MHC complexes and yet is often highly specific for a particular peptide–MHC combination¹.

The ability of a peptide–MHC complex to activate a T cell generally correlates with the strength and duration of TCR binding². We evaluated the contribution of individual residues in MCC-IE^k to TCR binding by mutating amino acids to alanine (or glycine if the original amino acid was alanine) and measuring the effects using surface plasmon resonance (Fig. 1 and Table 1). When mapped onto the structure of MCC-IE^k (ref. 4), these measurements delineate a broad contact surface for the 2B4 TCR (Fig. 2a), which is similar to that seen in the crystal structures of other TCR and peptide–MHC complexes³. MHC residues spanning Gln 61 to Ala 72 on the α -

chain and Gln 64 to Thr 77 on the β -chain of MCC-IE^k all contribute to binding affinity. These residues include charged, polar and neutral residues (such as β -chain Asp 76, β -chain Thr 77, α -chain Ala 65), and most contribute only modestly to binding stability (two- to fourfold decreases in affinity). A small cluster of residues spanning positions 64–73 on the β -chain has a larger impact on stability (8- to 40-fold decreases in affinity) and may comprise an MHC anchor point for TCR recognition, similar to that seen in a complex of a TCR and a class I MHC molecule⁵.

Notably, alanine substitutions at each of the three TCR contact positions in the MCC peptide^{6,7} result in very large decreases in affinity, such that binding is not measurable (on the basis of BIAcore 3000 sensitivity, this corresponds to more than a 200-fold decrease in binding). As compared with MHC contact residues, the much larger contribution of peptide contact residues to complex stability indicates that the peptide is a hotspot for TCR binding.

In apparent contrast to these observations, alanine scanning mutagenesis of the 2C TCR, in conjunction with competitive enzyme-linked immunoabsorbent binding assays to either the agonist ligand SIYR-K^b or the alloreactive ligand QL9-L^d, suggests a smaller energetic role for those portions of the TCR that contact peptide than those observed here^{8,9}. But the 2C TCR forms a poor complementary fit to peptide when bound to K^b, as compared with other TCR/peptide–MHC complexes^{10,11}. Perhaps these few poor contacts must be overcome by increasing the relative energetic importance of MHC contacts for stable binding, thereby contribut-

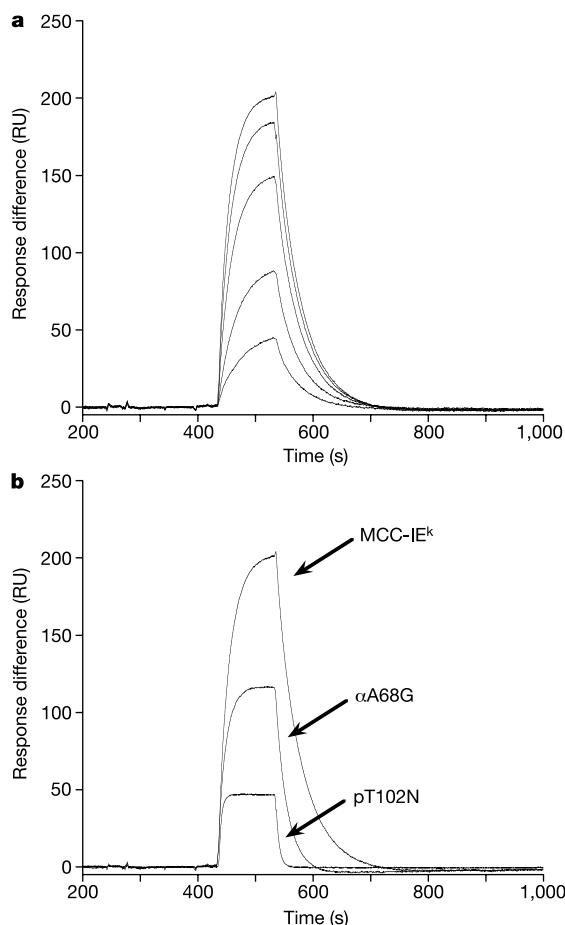


Figure 1 Binding of 2B4 TCR to wild-type MCC-IE^k and selected point mutants in the peptide–MHC epitope, monitored by surface plasmon resonance (BIAcore). RU, resonance units. **a**, Interaction of wild-type MCC-IE^k with 1, 2.5, 5, 7.5 and 10 μM 2B4 TCR. **b**, Representative binding curves for interaction of 2B4 TCR with wild-type MCC-IE^k, MHC α -chain mutant Ala68 → Gly and peptide mutant Thr102 → Asn.

ing to L^d alloreactivity. Alternatively, alanine scanning of 2C, whose CDR3 β loop consists largely of glycine, may miss important stabilizing contacts between the polypeptide backbone of the CDR3 β loop and the peptide ligand. Mutagenesis of peptide contacts for the 3L.2 (ref. 12) and A6 (ref. 13) TCRs suggests that a peptide energetic hotspot may be a more common feature of TCR and peptide–MHC interactions.

We analysed the contribution of individual residues in MCC-IE^k to the initial association with 2B4. The interactions that are important for this transition-state complex can be determined by analysing association rate constants. This type of analysis, phi (ϕ) value determination, has been used to study unimolecular protein folding pathways and the association-induced folding of protein fragments¹⁴. The ϕ value for a given residue incorporates its contribution to the activation energy of association, normalized by its contribution to the free energy of binding. Contacts with ϕ values of zero are not formed in the transition state nor involved in the initial association, whereas those with large ϕ values (up to a maximum of one, which represents full formation in the transition state) help to guide the course of the reaction.

In the 2B4/MCC-IE^k complex, 8 out of the 11 contacts on the MHC helices that contribute to binding affinity are also present in the transition state, with ϕ values ranging from 0.2 to 0.7 (Fig. 2b and Table 1). These residues comprise a continuous surface on the α - and β -chains of the MHC molecule and surround the central portion of the peptide antigen. ϕ values for the peptide contacts of 2B4 on MCC-IE^k could not be calculated from alanine mutations, because binding was undetectable. We therefore characterized a set of peptide variants in which portions of amino acid side chains were modified or removed without abolishing detectable TCR binding or grossly changing the peptide–MHC structure. By choosing peptide variants that in combination delineated each amino acid side chain, we could infer a ϕ value for each residue.

We found that TCR contacts with the MCC peptide mainly

influence the off rate of binding but not the on rate (Table 1). In particular, the ϕ values for all variants at positions 97 and 102 of the peptide are zero, which indicates that specific TCR contacts with these residues do not form in the transition state. The ϕ value for the central peptide contact, as judged by the Lys99 $\rightarrow$ Arg variant, is 0.17 and shows a high degree of specificity for the positive charge at the tip of the lysine side chain. This value is smaller than that for most of the MHC contacts and suggests only partial involvement in the transition state. Notably, the MCC-IE^k residues with the largest ϕ values are charged residues on the MHC helices. Molecular modelling indicated that these residues comprise patches of electrostatic complementarity with the 2B4 TCR, which is consistent with other studies that indicate that long-range electrostatic steering can have an important role in protein–protein association^{15,16}.

It is noteworthy that peptide contacts with TCR contribute substantially to complex stability and yet have little effect on association, whereas MHC contacts contribute little to stability but dominate association. This dichotomy is marked, as our measurements suggest that the MCC peptide accounts for $\sim$ 85% of the stabilization energy, but only $\sim$ 10% of the association energy (Methods). We conclude that MHC interactions guide the docking of the TCR with only minor contributions from the peptide, whereas peptide contacts largely determine the stability of binding.

To determine whether our observed functional subdivision of MCC-IE^k was a general property of peptide–MHC or a unique case, we analysed the data for another well-studied TCR/peptide–MHC system, A6/Tax-HLA-A2 (refs 5, 13). Although kinetic data for all of the residues comprising the A6 interaction surface on Tax-A2 were not available, we applied ϕ analysis to all published MHC and peptide contacts (see Supplementary Information). The available data suggest that MHC contacts are more prominent in the transition state than are peptide contacts. Four out of the six MHC contacts that we analysed have ϕ values between 0.5 and 1.0, whereas all peptide contacts have ϕ values that are less than 0.3

Table 1 Binding parameters for the 2B4 TCR-binding epitope on MCC-IE^k

	k_{on} (M ⁻¹ s ⁻¹)	k_{off} (s ⁻¹)	K_{eq} (μ M)	$K_{eq,mu}/K_{eq,wt}$	$\Delta\Delta G$ (kcal mol ⁻¹)	ϕ
Wild type	4,000 $\pm$ 370	0.022 $\pm$ 0.0003	5.5 $\pm$ 0.5			
MHC mutants						
α Q61A	3,920 $\pm$ 170	0.0336 $\pm$ 0.0001	8.6 $\pm$ 0.3	1.6 $\pm$ 0.2	0.278 $\pm$ 0.035	0.045*
α A65G	2,600 $\pm$ 190	0.0644 $\pm$ 0.0013	24.8 $\pm$ 2.0	4.5 $\pm$ 0.5	0.890 $\pm$ 0.099	0.29
α A68G	3,200 $\pm$ 300	0.0492 $\pm$ 0.002	15.4 $\pm$ 1.6	2.8 $\pm$ 0.4	0.610 $\pm$ 0.087	0.22
α K71A	2,020 $\pm$ 120	0.0286 $\pm$ 0.0003	14.3 $\pm$ 0.9	2.6 $\pm$ 0.3	0.566 $\pm$ 0.065	0.73
α A72G	2,780 $\pm$ 220	0.0638 $\pm$ 0.0031	22.9 $\pm$ 2.1	4.2 $\pm$ 0.5	0.850 $\pm$ 0.101	0.25
α E79A	3,960 $\pm$ 240	0.0217 $\pm$ 0.0009	5.5 $\pm$ 0.4	1.0 $\pm$ 0.1	0	0
β E59A	3,950 $\pm$ 630	0.022 $\pm$ 0.001	5.5 $\pm$ 0.9	1.0 $\pm$ 0.2	0	0
β S63A	3,910 $\pm$ 650	0.0225 $\pm$ 0.0014	5.75 $\pm$ 1.0	1.1 $\pm$ 0.2	0.056 $\pm$ 0.010*	0.076*
β Q64A	3,800 $\pm$ 420	0.41 $\pm$ 0.016	108 $\pm$ 13	20 $\pm$ 3	1.77 $\pm$ 0.27	0.017*
β E66A	1,150 $\pm$ 180	0.049 $\pm$ 0.002	42.6 $\pm$ 6.9	7.8 $\pm$ 1.4	1.22 $\pm$ 0.22	0.61
β E69A	2,100 $\pm$ 280	0.25 $\pm$ 0.005	119 $\pm$ 16	22 $\pm$ 4	1.83 $\pm$ 0.33	0.21
β A73G	1,280 $\pm$ 100	0.29 $\pm$ 0.01	227 $\pm$ 19	41 $\pm$ 5	2.20 $\pm$ 0.27	0.31
β D76A	2,110 $\pm$ 200	0.0498 $\pm$ 0.0017	23.6 $\pm$ 2.4	4.3 $\pm$ 0.6	0.864 $\pm$ 0.121	0.44
β T77A	3,900 $\pm$ 520	0.076 $\pm$ 0.0008	19.5 $\pm$ 2.6	3.6 $\pm$ 0.6	0.758 $\pm$ 0.126	0.02*
β H81A	4,000 $\pm$ 650	0.024 $\pm$ 0.006	6.0 $\pm$ 1.8	1.1 $\pm$ 0.3	0.056 $\pm$ 0.010*	0
α K71A/ β E66A	640 $\pm$ 10	0.045 $\pm$ 0.003	70 $\pm$ 5	13 $\pm$ 1.5	1.52 $\pm$ 0.18	0.72
Peptide mutants						
pY97A	ND	ND	ND	>200	>3.14	ND
pY97F	4,000 $\pm$ 190	0.0263 $\pm$ 0.0002	6.6 $\pm$ 0.3	1.2 $\pm$ 0.1	0.108 $\pm$ 0.009	0
pY97Cha	4,000 $\pm$ 260	0.112 $\pm$ 0.005	28 $\pm$ 2.2	5.1 $\pm$ 0.6	0.965 $\pm$ 0.114	0
pK99A	ND	ND	ND	>200	>3.14	ND
pK99Orn	ND	ND	ND	>200	>3.14	ND
pK99K(Ac)	ND	ND	ND	>200	>3.14	ND
pK99Nle	ND	ND	ND	>200	>3.14	ND
pK99R	2,070 $\pm$ 190	0.50 $\pm$ 0.04	240 $\pm$ 30	44 $\pm$ 7	2.24 $\pm$ 0.36	0.17
pT102A	ND	ND	ND	>200	>3.14	ND
pT102S	3,960 $\pm$ 160	0.134 $\pm$ 0.001	33.8 $\pm$ 1.4	6.1 $\pm$ 0.6	1.07 $\pm$ 0.11	0.006*
pT102N	4,030 $\pm$ 680	0.175 $\pm$ 0.005	43.4 $\pm$ 7.4	7.9 $\pm$ 1.5	1.22 $\pm$ 0.23	-0.004
pY97F/pT102S	4,200 $\pm$ 230	0.18 $\pm$ 0.003	42.9 $\pm$ 2.5	7.8 $\pm$ 0.8	1.22 $\pm$ 0.13	-0.024*

*Values are not significant because of the magnitude of error in the kinetic rate constants. These residues are therefore not considered to contribute to stability or association. ND, not determined. Cha, cyclohexylalanine; Orn, ornithine; K(Ac), acetyl-lysine; Nle, norleucine.

except for the central tyrosine, which protrudes into the CDR3 loops of the A6 TCR and has a ϕ value of 0.56. Given the unusually fast association rate of the A6 TCR ($10^5 \text{ M}^{-1} \text{ s}^{-1}$), as compared with most TCR agonist ligands (10^3 – $10^4 \text{ M}^{-1} \text{ s}^{-1}$; ref. 2), the ϕ values for its association might be expected to be higher throughout the epitope, as is observed.

We also measured the average backbone distances between TCR CDR loops and their contacts on MHC helices and peptide in the crystal structures of four TCR/peptide–MHC complexes, including two class-I-restricted and two class-II-restricted TCRs (Protein Data Bank (PDB) accession codes, 2CKB, 1FO0, 1D9K and 1FYT). We found that on average a greater distance separates the TCR from peptide than from MHC, especially for the CDR1 and CDR2 contacts with peptide–MHC. These differences range from 0.2 to 0.8 Å and, although small, are significant and show that an approaching TCR will initially encounter the MHC helices, which will present a structural barrier to the TCR and force longer range contacts with the peptide. Although the depth of the bound peptide varies in different class I and class II MHC molecules, with some class I peptides having an upward bulge, in general most of the peptide surface in peptide–MHC structures is deep in the binding groove and lies below the surfaces of the MHC helices³.

These structural features of TCR/peptide–MHC complexes, together with our current data, suggest a two-step mechanism for TCR recognition (Fig. 3). In this model, initial TCR–MHC interactions, aided by minor contributions from TCR contacts with the peptide, guide the TCR to its ligand. This is followed by a final folding of the two CDR3 loops of the TCR over the peptide. T-cell activation would be triggered only on the formation of stable peptide contacts. This model suggests that MHC restriction may be viewed as a ‘kinetic gatekeeper’ that promotes peptide sampling by guiding the docking of TCR onto peptide–MHC in a way that is mainly independent of peptide. It is intriguing that the CDR3 loops of TCRs are highly flexible and mobile in the unbound state¹⁷ and adopt conformations that cannot specifically contact peptide residues without substantial rearrangement, sometimes up to 15 Å (refs 3, 18). In contrast, the CDR1 and CDR2 loops are much more rigid, generally showing little or no rearrangement on binding peptide–MHC. Thus, the TCR may scan MHC molecules using a ‘lock and key’ type of binding with its CDR1 and CDR2 loops, followed by an

induced fit of its CDR3 loops over the peptide^{19,20}.

A two-step TCR recognition mechanism might be necessary to scan efficiently the many thousands of peptides displayed by MHC molecules on an antigen-presenting cell. It would also help explain the inherent crossreactivity of TCRs²¹, because the CDR3 loops of the TCR could adopt several conformations and thus form stable contacts with structurally very different peptides. Studies of other protein–protein interactions indicate that small subsections of a binding epitope can account for a significant portion of the total stabilization energy, comprising a binding ‘hotspot’²². In our studies, we have found that different parts of a binding surface may have distinct roles in the association and stabilization of a protein complex, comprising association and stabilization hotspots. This functional subdivision of a binding epitope might be especially important for interactions in which one binding partner must search quickly among many similar targets for an optimal one, such as in the recognition of specific sequence motifs by DNA-binding proteins.

The functional subdivision of the peptide–MHC epitope would have broad implications for T cell biology. One example is in thymic development, during which the TCR repertoire undergoes selection to yield T cells that are skewed to recognize self-MHC alleles^{23,24}. A study has indicated that MHC reactivity is inherent in the germline-encoded portions of the TCR²⁵, which suggests that most TCRs may have rudimentary scanning capability. Our studies indicate that the TCR recognition of MHC and peptide can be separable events, which suggests that thymic selection may imprint MHC restriction by enriching for TCRs that can scan a particular self-MHC and may be unable to scan other MHCs efficiently. Although most MHC polymorphisms are buried in the peptide-binding groove, high-resolution structural studies indicate that they can still have significant effects on MHC helix conformation and may thereby influence scanning^{26,27}. The mature TCR repertoire need not be enriched for specific V regions, however, because structural studies indicate that there is flexibility in MHC recognition by a given V region and therefore no specific MHC ‘footprint’ for TCR binding^{3,18}. Thymic

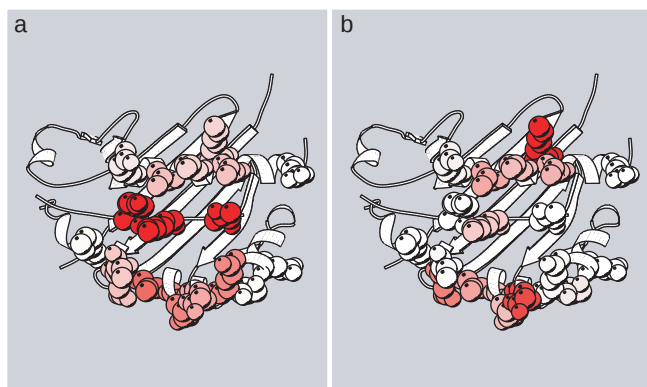


Figure 2 Structures of MCC-IE^K, showing the amino acid residues mutated in this study and their effects on TCR complex stability and association. The peptide-binding pocket of IE^K and the MCC peptide are shown as white ribbons. Residues studied are shown as space-filling representations and are colour-coded with respect to their effect on stability and association. **a**, Contributions of residues to complex stability ($\Delta\Delta G$). Residues are shaded white (no effect, $\Delta\Delta G = 0 \text{ kcal mol}^{-1}$) to red (maximum effect in this study, $\Delta\Delta G > 3.14 \text{ kcal mol}^{-1}$). **b**, Contributions of residues to complex association (ϕ). Residues are shaded white (no interaction, $\phi = 0$) to red (maximum interaction in this study, $\phi = 0.73$). Models were generated with MOLSCRIPT³⁰ using coordinates from the crystal structure of MCC-IE^K (ref. 4).

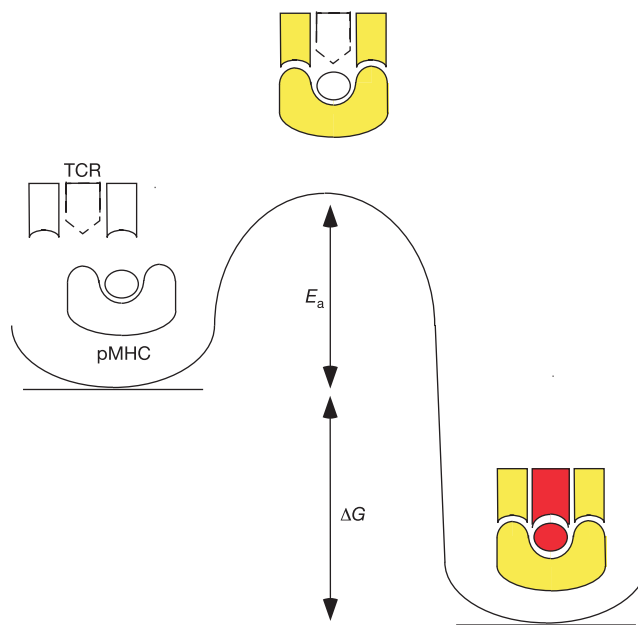


Figure 3 Model for TCR sampling of peptide–MHC (pMHC), showing formation of specific MHC and peptide contacts during the course of the binding reaction. The CDR1 and CDR2 loops of the TCR are primed to scan the MHC scaffolding, and form specific contacts (yellow) in the transition state that guide the TCR to a general orientation over peptide. The CDR3 loops show structural plasticity and form strong specific peptide contacts (red) by induced-fit binding subsequent to the docking of TCR on MHC.

selection may also be assisted by the inherent crossreactivity of the TCR that arises from CDR3 loop flexibility, which enables interactions with self-peptide-MHC that may resemble a foreign peptide-MHC²⁸ or may be very different. A similar process may be involved in peripheral T-cell homeostasis, which also depends on recognition of self-peptide-MHC²⁹. □

Methods

Alanine-scanning mutagenesis

Sites of mutation were chosen on the basis of the following two criteria: first, positions known from T-cell activation studies using peptide-MHC mutants to have biological effects on T-cell recognition (refs 6, 7; and M. Krosgaard and L.C.W., unpublished data); and second, likely positions of contact deduced from homology modelling using the crystal structures of 2C/dEV8-H2K^b (PDB accession code 2CKB) and IE^k bound to a haemoglobin peptide (PDB accession code 1IEA). We chose residues such that mutations would affect only interaction with the TCR and not the structure of the rest of the peptide-MHC surface. Notably, the three peptide residues that form TCR contacts point up and away from the peptide-binding cleft and do not contact MHC. In addition, we minimized complications arising from conformational heterogeneity by mutating residues on the peptide-MHC, which undergoes minimal conformational change on binding TCR. There was no evidence for large-scale destabilization or unfolding of peptide-MHC owing to our mutations, as all mutants contained bound peptide, had small and distinct effects on association and dissociation, and were bound by an antibody specific for native structure (14.4.4). Local structural perturbations caused by mutations could affect this type of analysis by changing the effective concentration of molecules that are competent for binding; however, the rate of such conformational fluctuations is much faster than our observed protein association rates, suggesting a pre-equilibrium of binding-competent species. In combination, our mutants comprised about 75% of the total surface area buried under the footprint of the TCR and covered the highest contact points on the peptide-MHC surface.

Protein production and purification

We produced MCC-IE^k and the various mutants by *in vitro* refolding of bacterially produced recombinant α - and β -subunits with chemically synthesized peptides. MHC mutants were produced by PCR site-directed mutagenesis and verified by DNA sequencing. The α -subunit of IE^k contained a carboxy-terminal amino acid tag encoding a signal sequence for attaching biotin with the BirA enzyme. Bacterial IE^k subunits were expressed in *Escherichia coli* BL21 DE3 pLysS as inclusion bodies. After bacterial lysis, inclusion bodies were washed with detergent buffers. Subunits were solubilized and oxidized in 6 M guanidinium hydrochloride and then diluted to a final concentration of 2 μ M in a redox folding buffer containing 50 mM sodium phosphate (pH 7.5), 25% glycerol, 2 mM EDTA, 5 mM reduced and 0.5 mM oxidized glutathione, 10 μ M peptide (MCC 88–103 wild-type sequence, ANERADLIAYLKQATK) and protease inhibitors. Correctly folded material was purified by affinity chromatography with 14.4.4. antibody against IE^k, biotinylated with BirA enzyme, and size-purified on an S200 sephacryl fast protein liquid chromatography (FPLC) sizing column.

We produced 2B4 TCR as a glycosyl phosphatidylinositol (GPI)-linked membrane-bound protein in transfected Chinese hamster ovary (CHO) cells using a Cell Pharm system (Cellex Biosciences). Protein was cleaved from the CHO cells with PI-PLC enzyme and purified sequentially on two antibody affinity columns (A2B4, against TCR α and H57, against TCR β). TCR was size-purified on an S200 sephacryl FPLC column before use.

BIAcore measurements of binding

Kinetic parameters for 2B4 binding to wild-type MCC-IE^k and mutants were obtained on BIAcore 2000 and 3000 instruments (BIAcore). Peptide-MHC was attached by the C-terminal biotin tag on the IE^k α -subunit to a streptavidin sensor chip surface to a total resonance unit (RU) value of 250–500. Binding constants were not dependent on the amount of peptide-MHC coupled to the sensor chip. Subsequent to sensor chip loading with peptide-MHC, extra biotin-binding sites were blocked by injection of free biotin. All measurements were baseline-corrected by subtracting TCR injected over a blank, biotin-blocked surface, and all mutant measurements were referenced to wild-type MCC-IE^k, which was measured simultaneously. We injected 2B4 TCR in PBS plus 0.005% P20 (BIAcore) at a flow rate of 30 μ l min⁻¹ at 25 °C, over a concentration range of 1–150 μ M according to the peptide-MHC mutant being measured. Data were fit to a 1:1 Langmuir binding model using both separate on and off rate determinations and global fitting (BiaEval, BIAcore). Rate constants were the same, within error, for both methods of fitting and did not vary systematically with the concentration of 2B4 TCR. Two residues (Glu 79 on the α -chain and Glu 59 on the β -chain of IE^k), which are located far outside the predicted 2B4 binding surface and do not affect T-cell activation⁶, do not contribute to 2B4 binding affinity. As predicted, the kinetic binding parameters of alanine mutants at these positions are identical to those of wild-type MCC-IE^k.

Data analysis

The dissociation of 2B4 from MCC-IE^k is described by the following equation:

$$[2B4/MCC-IE^k] \xrightleftharpoons[k_{on}]{k_{off}} [2B4] + [MCC-IE^k]$$

where k_{off} is the dissociation constant and k_{on} is the association constant. R is the gas constant and T is the temperature in Kelvin. The equilibrium constant for dissociation is $K_{eq} = k_{off}/k_{on}$, and $\Delta G = RT \ln K_{eq}$. The association constant for binding can be

converted to an activation energy, $E_a = -RT \ln k_{on}$. Comparison of activation energies for wild-type versus mutant MCC-IE^k are given as ϕ values, which are calculated as ΔE_a normalized by $\Delta \Delta G$ (ref. 14):

$$\phi = -RT \ln(k_{on,wt}/k_{on,mu}) / -RT \ln(K_{eq,mu}/K_{eq,wt})$$

To estimate the contribution of peptide to stabilization and activation energies, we first calculated a total stabilization energy of 7.17 kcal mol⁻¹ and a total activation energy of 4.91 kcal mol⁻¹, which were based on an equilibrium binding constant of 5.5 μ M and an association rate of 4,000 M⁻¹ s⁻¹, for the binding of 2B4 to wild-type MCC-IE^k.

For stabilization, we assumed that the MCC peptide was subdivided into two largely independent contact points comprising residues Tyr 97/Lys 99 and Thr 102. This was based on structural studies of MCC-IE^k, which indicate that Tyr 97 and Lys 99 are in close proximity⁴, and on binding measurements of the double peptide mutant Tyr97Phe/Thr102Ser, which show the additive effects of the single mutants. Assuming that each peptide contact point accounts for 3.14 kcal mol⁻¹ (on the basis of alanine mutations at each of the three peptide contact positions), we calculated a net peptide contribution that is about 85% of the total binding energy.

For association, only position 99 of the peptide contributes to the activation energy. Assuming that Lys 99 contributes a stabilization energy of 3.14 kcal mol⁻¹, its ϕ value of 0.17 yields a net peptide contribution of 0.53 kcal mol⁻¹, which is about 10% of the total activation energy.

TCR/peptide-MHC distance measurements

We measured TCR/peptide-MHC distances with the program O7, using lists of published TCR/peptide-MHC contacts (both van der Waals and hydrogen-bond contacts). Distances were measured between the backbone α -carbons of contacting residues, which were classified as either TCR-peptide or TCR-MHC, and averaged for comparison.

Received 19 February; accepted 24 May 2002; doi:10.1038/nature00920.

1. Germain, R. N. & Stefanova, I. The dynamics of T cell receptor signalling: complex orchestration and the key roles of tempo and cooperation. *Annu. Rev. Immunol.* **17**, 467–522 (1999).
2. Davis, M. M. *et al.* Ligand recognition by $\alpha\beta$ T cell receptors. *Annu. Rev. Immunol.* **16**, 523–544 (1998).
3. Rudolph, M. G. & Wilson, I. A. The specificity of TCR/pMHC interaction. *Curr. Opin. Immunol.* **14**, 52–65 (2002).
4. Fremont, D. H. *et al.* Structural basis of cytochrome c presentation by IE^k. *J. Exp. Med.* **195**, 1043–1052 (2002).
5. Baker, B. M., Turner, R. V., Gagnon, S. J., Wiley, D. C. & Biddison, W. E. Identification of a crucial energetic footprint on the α 1 helix of human histocompatibility leukocyte antigen (HLA)-A2 that provides functional interactions for recognition by Tax peptide/HLA-A2-specific T cell receptors. *J. Exp. Med.* **193**, 551–562 (2001).
6. Ehrlich, E. W. *et al.* T cell receptor interactions with peptide/major histocompatibility complex (MHC) and superantigen/MHC ligands is dominated by antigen. *J. Exp. Med.* **178**, 713–722 (1993).
7. Reay, P., Kantor, R. M. & Davis, M. M. Use of global amino acid replacements to define the requirements for MHC binding and T cell recognition of moth cytochrome c (93–103). *J. Immunol.* **152**, 3946–3957 (1994).
8. Manning, T. C. *et al.* Alanine scanning mutagenesis of an $\alpha\beta$ T cell receptor: mapping the energy of antigen recognition. *Immunity* **8**, 413–425 (1998).
9. Lee, P. U., Churchill, H. R., Daniels, M., Jameson, S. C. & Kranz, D. M. Role of 2C T cell receptor residues in the binding of self- and allo-major histocompatibility complexes. *J. Exp. Med.* **191**, 1355–1364 (2000).
10. Degano, M. *et al.* A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. *Immunity* **12**, 251–261 (2000).
11. Garcia, K. C. *et al.* Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* **279**, 1166–1172 (1998).
12. Kersh, G. J., Kersh, E. N., Fremont, D. H. & Allen, P. M. High- and low-potency ligands with similar affinities for the TCR: the importance of kinetics in TCR signaling. *Immunity* **9**, 817–826 (1998).
13. Baker, B. M., Ding, Y.-H., Garboczi, D. N., Biddison, W. E. & Wiley, D. C. Structural, biochemical, and biophysical studies of HLA-A2/alter peptide ligands binding to viral-peptide-specific human T-cell receptors. *Cold Spring Harb. Symp. Quant. Biol.* **64**, 235–241 (1999).
14. Fersht, A. R. Nucleation steps in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3–9 (1997).
15. Schreiber, G. & Fersht, A. R. Rapid electrostatically assisted association of proteins. *Nature Struct. Biol.* **3**, 427–431 (1996).
16. Davis, S. J., Davies, E. A., Tucknott, M. G., Jones, E. Y. & van der Merwe, P. A. The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proc. Natl Acad. Sci. USA* **95**, 5490–5494 (1998).
17. Hare, B. J. *et al.* Structure, specificity, and CDR mobility of a class II restricted single-chain T cell receptor. *Nature Struct. Biol.* **6**, 574–581 (1999).
18. Reiser, J. B. *et al.* AT cell receptor CDR β loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. *Immunity* **16**, 345–354 (2002).
19. Boniface, J. J., Reich, Z., Lyons, D. S. & Davis, M. M. Thermodynamics of T cell receptor binding to peptide-MHC: evidence for a general mechanism of molecular scanning. *Proc. Natl Acad. Sci. USA* **96**, 11446–11451 (1999).
20. Willcox, B. E. *et al.* TCR binding to peptide-MHC stabilizes a flexible recognition interface. *Immunity* **10**, 357–365 (1999).
21. Mason, D. A very high level of cross-reactivity is an essential feature of the T-cell receptor. *Immunol. Today* **19**, 395–404 (1998).
22. Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383–386 (1995).
23. Bevan, M. J. In a radiation chimera, host H-2 antigens determine immune responsiveness of donor cytotoxic cells. *Nature* **269**, 417–418 (1977).
24. Zinkernagel, R. M. *et al.* On the thymus in the differentiation of 'H-2 self-recognition' by T cells: evidence for dual recognition? *J. Exp. Med.* **147**, 882–896 (1978).
25. Zerrahn, J., Held, W. & Raulet, D. H. The MHC reactivity of the T cell repertoire prior to positive and negative selection. *Cell* **88**, 627–636 (1997).

26. Hennecke, J. & Wiley, D. C. Structure of a complex of the human α/β T cell receptor (TCR) HA1. (7), influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. *J. Exp. Med.* **195**, 571–581 (2002).
27. Luz, J. G. *et al.* Structural comparison of allogeneic and syngeneic T cell receptor-peptide-major histocompatibility complex complexes: a buried alloreactive mutation subtly alters peptide presentation substantially increasing V β interactions. *J. Exp. Med.* **195**, 1175–1186 (2002).
28. Savage, P. A. & Davis, M. M. A kinetic window constricts the T cell receptor repertoire in the thymus. *Immunity* **14**, 243–252 (2001).
29. Tanchot, C., Lemonnier, F. A., Pérarnau, B., Freitas, A. A. & Rocha, B. Differential requirements for survival and proliferation of CD8 naive or memory T cells. *Science* **276**, 2057–2058 (1997).
30. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. Boniface, C. Gerke, S. Hedrick, D. Herschlag, J. Huppa, M. Krogsgaard, M. Kuhns, Z. Reich, R. Sciammas and members of the Davis lab for discussions and comments on the manuscript; B. Malissen and D. Fremont for communicating results before publication; N. Prado for assistance with TCR production; and A. Cochran and H. Lowman for use of their BIAcore 2000. L.C.W. is a postdoctoral fellow of the Cancer Research Fund of the Damon Runyon Walter Winchell Foundation. D.S.L. was a Howard Hughes Medical Institute predoctoral fellow. This work was supported by the NIH (M.M.D. and K.C.G.), the Howard Hughes Medical Institute (M.M.D.), the Multiple Sclerosis Society (K.C.G.), and the Cancer Research Institute (K.C.G.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M.D. (e-mail: mdavis@cmgm.stanford.edu).

APC-dependent proteolysis of the mitotic cyclin Clb2 is essential for mitotic exit

Ralph Wäsch & Frederick R. Cross

The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Cyclin degradation is central to regulation of the cell cycle. Mitotic exit was proposed to require degradation of the S phase cyclin Clb5 by the anaphase-promoting complex^{1,2} activated by Cdc20 (APC^{Cdc20})³. Furthermore, Clb5 degradation was thought to be necessary for effective dephosphorylation and activation of the APC regulatory subunit Cdh1 (also known as Hct1) and the cyclin-dependent kinase inhibitor Sic1 by the phosphatase Cdc14, allowing mitotic kinase inactivation and mitotic exit^{3–7}. Here we show, however, that spindle disassembly and cell division occur without significant APC^{Cdc20}-mediated Clb5 degradation, as well as in the absence of both Cdh1 and Sic1. We find instead that destruction-box-dependent degradation of the mitotic cyclin Clb2 is essential for mitotic exit. APC^{Cdc20} may be required for an essential early phase of Clb2 degradation, and this phase may be sufficient for most aspects of mitotic exit. Cdh1 and Sic1 may be required for further inactivation of Clb2-Cdk1, regulating cell size and the length of G1.

CDC20 is essential for mitosis. Cdc20-dependent degradation of the securin Pds1 is required for chromosome separation. *pds1* deletion (*pds1Δ*) allows chromosome separation but not mitotic exit in *cdc20Δ* cells⁸, suggesting that additional APC^{Cdc20} targets must be degraded. Cdc20-dependent degradation of Clb5 was proposed to be essential for mitotic exit, as *cdc20Δ pds1Δ clb5Δ* cells are viable³.

To determine whether APC^{Cdc20}-dependent Clb5 degradation was dependent on the Clb5 destruction box (D box)^{9,10} we expressed Clb5 and Clb5Δdb (where the D box is deleted) from the *GAL1* promoter in *cdc20Δ* or *CDC20* cells. *cdc20* deletion stabilized Clb5 (ref. 3). Deletion of the Clb5 D box stabilized Clb5 similarly with or without *CDC20* (Fig. 1a), suggesting that the D box was the main target for Cdc20 recognition. Inactivation of additional candidate Clb5 D boxes (RXXL), SCF or Cdh1 did not affect Clb5 stability (Supplementary Information Fig. S1a, b). Unexpectedly, D-box deletion from endogenous *CLB5* had little cell cycle phenotype, although Clb5 abundance was significantly deregulated (Figs 1b, c and 2f, g). Thus D-box-dependent Clb5 degradation is not required for exit from mitosis and viability.

To examine the D-box dependence of Clb5 proteolysis at endogenous expression levels during mitotic exit, we constructed *cdc20Δ pds1Δ GAL-CDC20* yeast strains³ containing *CLB5* or *CLB5Δdb* (exact gene replacement). Both strains arrested before mitotic exit on glucose medium (unlike *cdc20Δ pds1Δ clb5Δ* (ref. 3)), indicating that *CLB5Δdb* restrained mitotic exit in the *cdc20Δ pds1Δ* background. Therefore it is unlikely that the Clb5 D box is required for inhibition of mitotic exit, although it could still be required for other untested mitotic Clb5 functions. *GALL-CDC20* re-induction released mitotic arrest (Fig. 1d–f), and wild-type Clb5 was completely degraded 20 min later. Clb5Δdb showed only moderate decline at later time points. The mitotic cyclins Clb2 and Clb3 were degraded rapidly after release, in both strains, and this degradation initiated while Clb5Δdb levels were high (Fig. 1e, f). Thus, significantly inhibiting APC^{Cdc20}-dependent Clb5 degradation had no effect on mitotic exit (Fig. 1d–f).

The moderate decline in levels of Clb5Δdb in this experiment could result from minor Cdc20-dependent degradation signals in Clb5 that are independent of the D box. This decline is probably not required for mitotic exit, as strong overexpression of Clb5Δdb-haemagglutinin (HA) from the *GAL1* promoter caused lethality but did not block mitotic exit¹⁰. Furthermore, amino-terminal Myc-tagging made Clb5Δdb almost completely stable, and was lethal at endogenous expression levels without an evident block to mitotic exit (Supplementary Information Fig. S2a, b). It is possible that the moderate Cdc20-dependent decline in Clb5Δdb could combine with inhibition by the Sic1 Clb kinase inhibitor¹¹ to effectively eliminate Clb5Δdb activity at endogenous expression levels. Therefore, to more fully deregulate Clb5Δdb activity we also deleted *SIC1*. *CLB5Δdb sic1Δ GAL-SIC1* strains were semi-lethal on glucose medium (*GAL-SIC1* off) (Fig. 2f), but a shift to glucose caused accumulation of cells with replicated but unseparated DNA, and without long mitotic spindles (Fig. 2b). Thus these cells were primarily defective in entry into mitosis and not in mitotic exit.

We constructed a *CLB5Δdb sic1Δ GAL-SIC1 cdc15-2* strain to examine a single cycle of mitotic exit. These cells exited mitosis (disassembled mitotic spindles) after release from *cdc15-2* anaphase block, despite Sic1 depletion (to below the level of detection) for 4 h in glucose medium at 37°C (Fig. 3a). Compare Sic1 re-accumulation in *cdc15-2 SIC1* strains, 10–20 min before spindle disassembly (Fig. 3a). Unlike Clb5, Clb5Δdb persisted through the *cdc15-2* anaphase block and subsequent release (Fig. 3a).

To examine mitotic exit in *sic1Δ CLB5Δdb* cells in an unperturbed cell cycle, we incubated a *sic1Δ CLB5Δdb GAL-SIC1* strain in glucose medium to deplete Sic1, elutriated small-budded cells and released into mating-factor to block the succeeding cell cycle after mitosis. Incubation in glucose for 30 min depleted Sic1 to about physiological levels in the initial elutriated fraction, but cell division occurred normally (Fig. 3b). Incubation in glucose for 60 min depleted Sic1 to below the level of detection, and most cells still completed cell division on schedule (Fig. 3b). The small number of cells that failed to divide were mostly blocked before nuclear division. Thus, we observed no inhibition of mitotic exit by Clb5Δdb, even in the absence of Sic1. Clb5Δdb only gradually

declined through these time courses, in contrast to the sharp drop in wild-type Clb5.

CLB5Δdb sic1Δ GAL-SIC1 cells incubated transiently in glucose showed high loss rates of a single-replication-origin plasmid but not of a multi-origin plasmid¹² (Fig. 2g), suggesting problems with origin usage¹⁰. Defects in chromosomal replication could explain reduced viability and pre-mitotic delay of *CLB5Δdb sic1Δ* mutants (Figs 2b, f and 3a, b).

Undegraded Clb5 was suggested to block mitotic exit by phosphorylation and inactivation of Sic1 and Cdh1 (refs 3, 4, 6). We found, however, that *sic1Δ cdh1Δ cdc15-2 GAL-SIC1* strains disassembled mitotic spindles on *cdc15-2* release after depletion of Sic1 (Fig. 3a). These cells also exhibit significant replication of DNA in the next cell cycle (Supplementary Information Fig. S3b). Furthermore, *sic1Δ cdh1Δ CDC15 GAL-SIC1* strains depleted of Sic1 could complete cell division after release of a nocodazole block (Supplementary Information Fig. S4). Shifting *cdh1Δ sic1Δ GAL-SIC1* strains from galactose to glucose results not in a block of mitotic exit, but in cell cycle delays with replicated but unseparated DNA without long mitotic spindles, similar to *CLB5Δdb sic1Δ GAL-SIC1* strains (Fig. 2b, d, e). Both strains show slow proliferation on glucose medium (Fig. 2f). An interesting difference that we cannot

yet interpret is endoreplication in *cdh1Δ sic1Δ* strains (Fig. 2d, e).

Clb2-associated H1 kinase activity does not drop during mitotic release of the *cdh1Δ sic1Δ cdc15-2* strain, unlike in wild type (Fig. 3a). Transient inactivation of mitotic kinase is sufficient to promote re-replication at different positions of the cell cycle, leading to the suggestion that inactivation of mitotic kinase at mitotic exit might be required for accurate DNA replication¹³. This may account for reduced viability of *cdh1Δ sic1Δ* strains and *CLB5Δdb sic1Δ* strains, and for plasmid loss in *CLB5Δdb sic1Δ* strains (Fig. 2f, g).

A 10-min delay in the loss of mitotic spindles on *cdc15* release in a *cdh1Δ* strain¹⁴ is also observed in the *cdh1Δ sic1Δ* double mutant (Fig. 3a). There are also some differences in microtubule morphology after mitosis in *cdh1Δ* and *cdh1Δ sic1Δ* strains compared with wild type (Supplementary Information Fig. S3a). These differences could be due to incomplete Clb degradation, or to a failure to degrade Ase1, a protein required for spindle maintenance. Ase1 degradation is Cdh1-dependent^{15,16}, and expression of undegradable Ase1 during *cdc15* release causes a 10-min delay in spindle breakdown¹⁵. Undegraded Clb5 was proposed to inactivate Cdh1 (refs 3, 6, 17, 18). Consistent with this, the Cdh1 target Clb2 is almost as stable on release of a *CLB5Δdb sic1Δ* strain as in a *cdh1Δ sic1Δ* strain (Fig. 3a). But occurrence of most aspects of mitotic exit

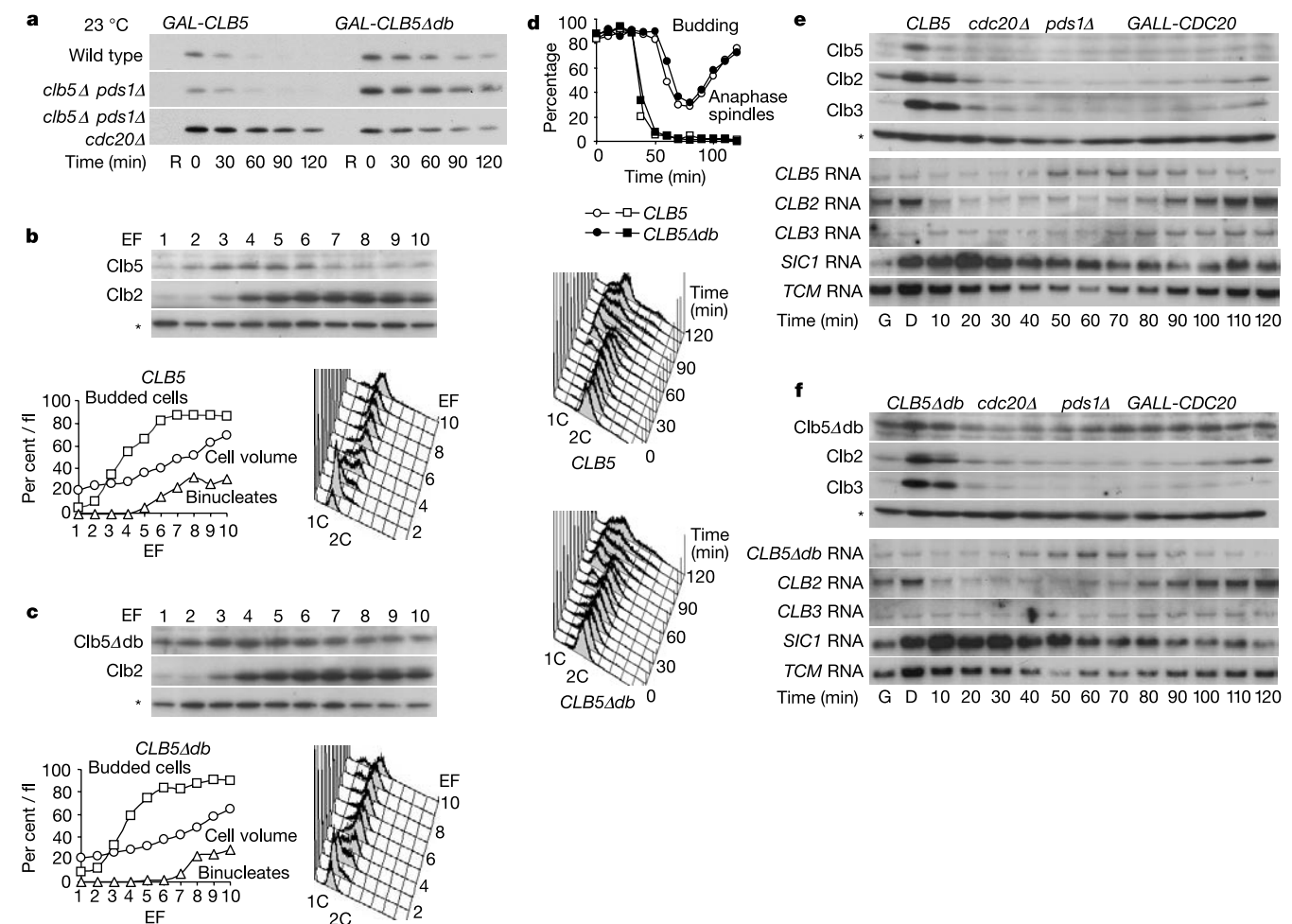


Figure 1 Clb5 degradation is Cdc20 and D-box dependent, but Clb5Δdb does not block mitotic exit. **a**, Stability of Clb5 and Clb5Δdb expressed from the galactose-inducible *GAL1* promoter, and then repressed, in the indicated backgrounds (see Supplementary Information for densitometric quantification). R, raffinose medium (uninduced control). **b**, **c**, Elutriation of *CLB5* (**b**) and *CLB5Δdb* strains (**c**). Immunoblots, budding (%) cell volume (in femtolitres, fl) and DNA content from elutriated fractions (EF) are shown.

d–f, *cdc20Δ pds1Δ GAL-SIC1* (*CLB5* or *CLB5Δdb*) cells were grown in YEPGal (G), blocked in YEPD (D, 4 h), and released from mitotic block in YEPGal. **d**, Budding (circles), anaphase spindles (squares) and DNA content (middle and bottom panels). **e**, **f**, Expression of *CLB5* (**e**) and *CLB5Δdb* (**f**) in the indicated backgrounds. Clb5, Clb2 and Clb3 immunoblots and *CLB5*, *CLB2*, *CLB3*, *SIC1* and *TCM1* RNA (as a loading control) are shown. Asterisks in **b**, **c**, **e**, **f** indicate nonspecific loading control.

in *CLB5Δdb sic1Δ* and *cdh1Δ sic1Δ* strains suggests that this level of Clb2 does not prevent mitotic exit, even without Sic1 inhibition.

The Cdc14 phosphatase is essential for mitotic exit, and Cdc14 dephosphorylates and activates Cdh1 and Sic1 (ref. 5). If these were the only Cdc14 targets, then *sic1Δ cdh1Δ* deletion should block mitotic exit. Thus Cdc14 probably has additional targets, such as Cdc15 (ref. 19) or Cdc6 (ref. 20). Cdc6 may be redundant with Sic1 for inhibition of Clb kinase in mitotic exit²⁰, and if it is activated by Cdc14, then this could allow for some regulation during mitotic exit even in the *sic1Δ cdh1Δ* strain. However, we did not detect inhibition in bulk Clb2-associated kinase measurements in *cdh1Δ sic1Δ* strains exiting mitosis (Fig. 3a). Perhaps the relevant inhibition occurs locally at origins of replication, where Cdc6 is bound.

Our data argue against the idea³ that Cdc20-dependent Clb5 degradation is required for mitotic exit. Neither blocking Clb5 degradation, nor removing Cdh1 and Sic1, the presumed targets of Clb5 (ref. 3), prevents mitotic exit. However, the role of Cdh1 in mitosis might be more complex, because other proteins important for mitotic control are subject to Cdh1-dependent degradation, including Ase1 (refs 15, 16), Cdc5 (ref. 21), and perhaps Cdc20 itself^{17,22}.

Some of the mitotic cyclins Clb1–4 might be essential APC^{Cdc20} targets. Although Cdh1 is significant in controlling proteolysis of the major mitotic cyclin Clb2 (refs 14, 16), Cdc20 may provide a

first phase of Clb2 degradation operating at higher levels of Clb2 (ref. 23).

To test the requirement for Clb2 degradation, we deleted the D box from the endogenous *CLB2* gene. *SIC1* overexpression was required for viability of *CLB2Δdb* strains (Fig. 2c, f). Shut-off of *GAL-SIC1* in *CLB2Δdb GAL-SIC1* strains caused arrest as large budded cells, with replicated DNA, separated nuclei and long mitotic spindles (Fig. 2c). These cells frequently exhibited multiple buds after prolonged arrest (not shown), but still most contained a single, long anaphase spindle. We were surprised that blocking Clb2 degradation by deletion of the D box prevented mitotic exit, because although Cdh1 was thought to control most Clb2 degradation, *cdh1* deletion does not block mitotic exit or viability^{14,16}. Also, *CLB2Δdb* expressed from a truncated promoter¹⁴ or the *CLB5* promoter⁹ allows viability. With precise endogenous gene replacement, however, *CLB2Δdb* (unlike *CLB5Δdb*) blocks mitotic exit.

Failure to exit mitosis owing to *CLB2Δdb* was also observed in *cdc15-2* block and release (Fig. 3a, and Supplementary Information Fig. S3a, b), and nocodazole block and release (Supplementary Information Fig. S4). This failure was associated with high levels of Clb2Δdb. Notably, more Clb2 accumulated in *CLB2Δdb* strains (failing mitotic exit) than in *cdh1Δ sic1Δ* strains (achieving mitotic exit) (Fig. 3a).

Clb2Δdb may block mitotic exit nearly autonomously. *clb5Δ*

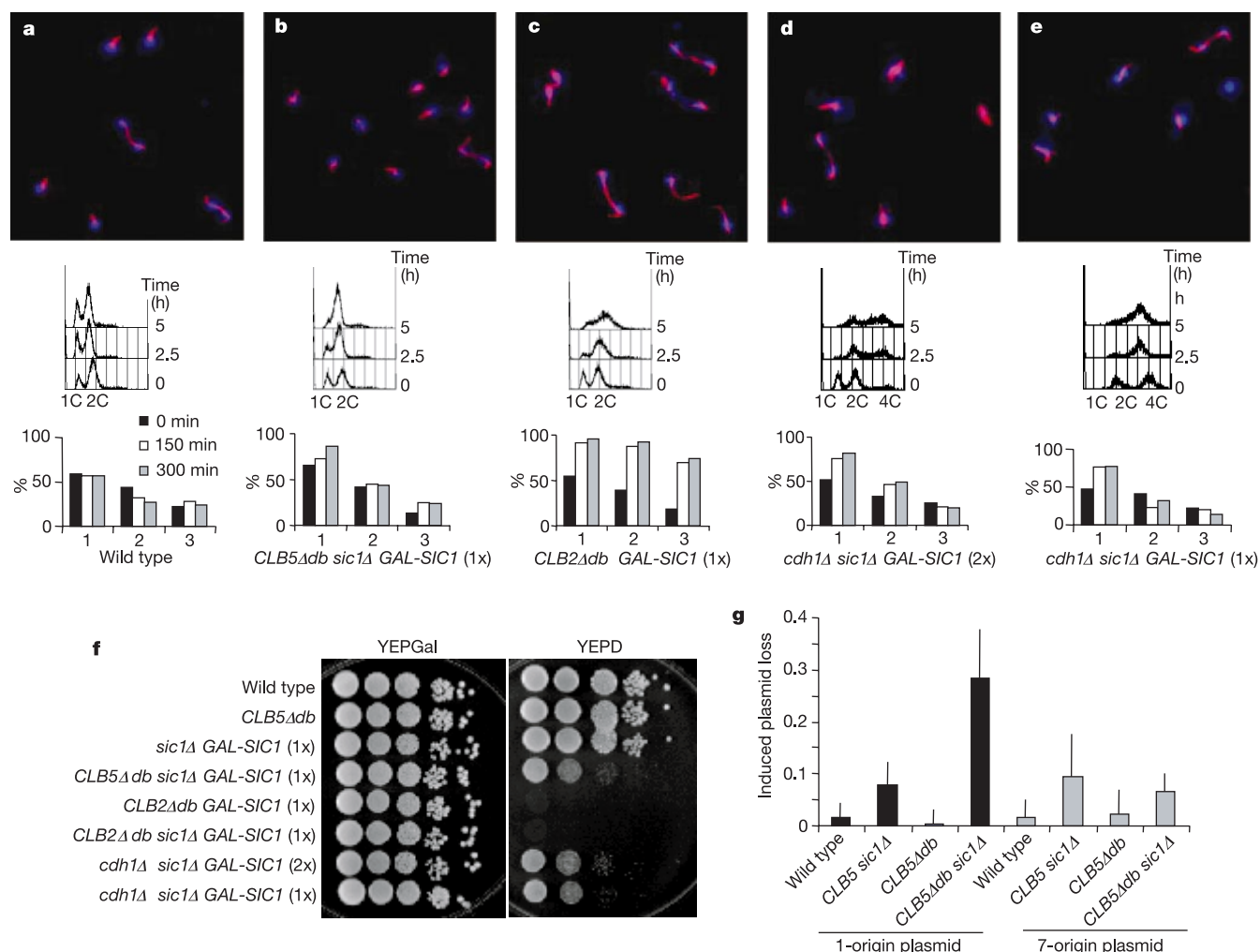


Figure 2 *CLB2Δdb* but not *CLB5Δdb sic1Δ* or *cdh1Δ sic1Δ* blocks mitotic exit.

a–e, Cells of the indicated genotypes stained for spindles (anti-tubulin; red) and DNA (DAPI; blue) after depletion of *GAL-SIC1* for 2.5 h. DNA content, percentage of budding (1), binucleates (2) and anaphase spindles (3) are also shown. (The *cdh1Δ sic1Δ*

GAL-SIC1 (1 ×) strain was diploidized; see Methods.) **f**, 1:10 serial dilutions of the indicated strains on galactose (*GAL-SIC1* on) or glucose (*GAL-SIC1* off), at 30 °C. **g**, The indicated *GAL-SIC1* strains containing single- and multiple-origin plasmids¹² were tested for plasmid loss after transient *GAL-SIC1* shut-off (see Methods).

CLB2Δdb GAL-SIC1 strains arrested on a shift to glucose medium similarly to *CLB5* controls (data not shown). Also, *clb1,3Δ CLB2Δdb GAL-SIC1* and *clb3,4Δ CLB2Δdb GAL-SIC1* strains were inviable on glucose medium (data not shown). (We could not construct a *clb1,3,4Δ CLB2Δdb GAL-SIC1* strain, for unknown reasons.)

If Clb2 is an essential target for Cdc20-mediated D-box-dependent degradation along with Pds1 (ref. 8), this could explain mitotic arrest of both *cdc20Δ pds1Δ* strains and *CDC20 PDS1 CLB2Δdb*

strains. Clb2 was strongly stabilized by either deletion of *cdc20* or deletion of the D box (Fig. 4a). *cdh1* deletion had less marked effects, so in cycling cells Cdc20 may control the bulk of D-box-dependent Clb2 degradation (Fig. 4c).

To examine Cdc20-dependent Clb2 degradation at endogenous Clb2 expression levels in synchronized cells, we blocked and released *cdc20Δ GAL-CDC20* strains carrying either *CDH1* or *cdh1Δ* (Fig. 4f). A significant proportion of Cdc20-dependent Clb2 degradation

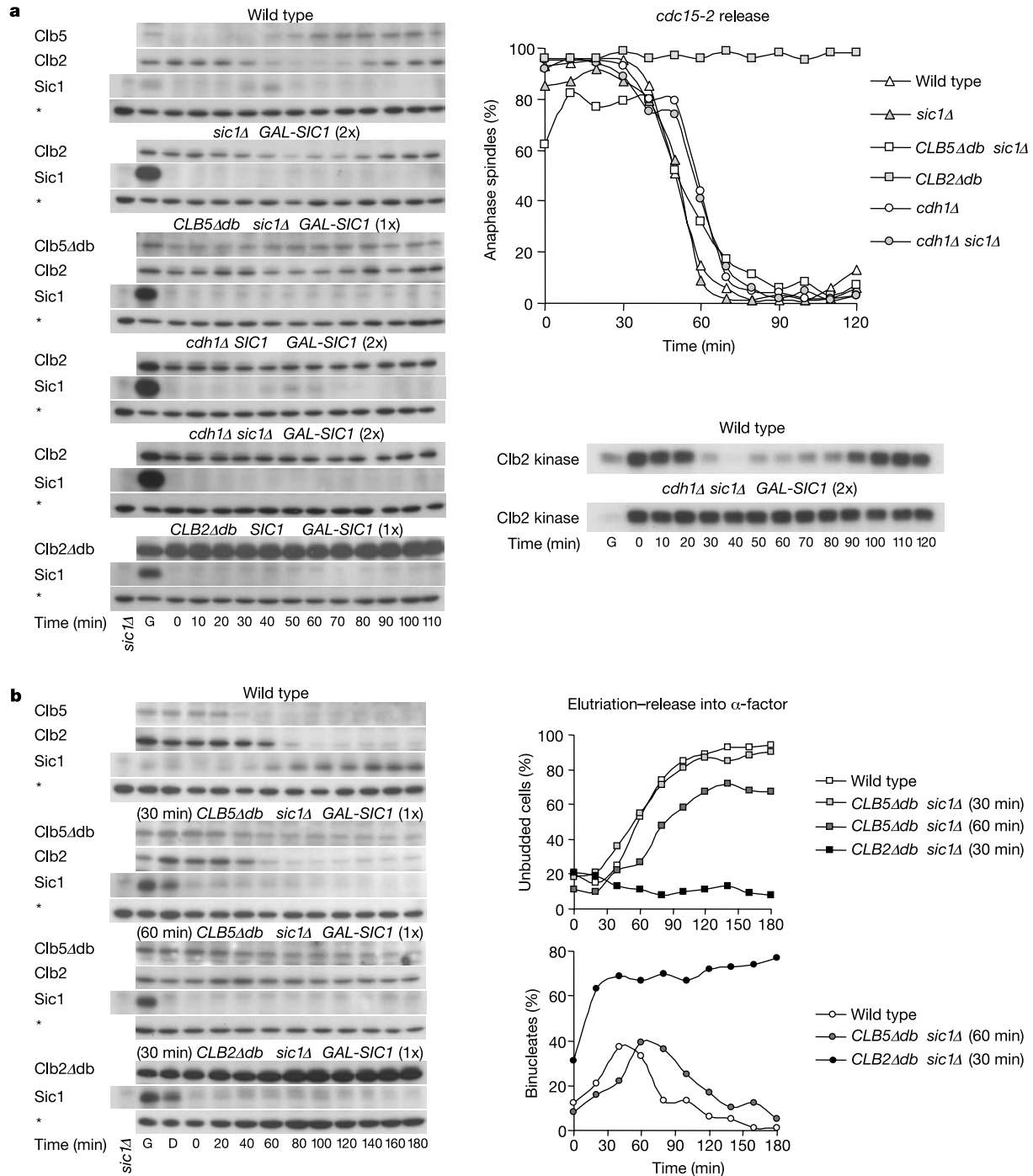


Figure 3 Kinetics of mitotic exit. **a**, *cdc15-2* strains with the indicated genotypes were grown in YEPGal (G) at 23°C, then transferred to YEPD (D) at 37°C for 4 h to deplete *GAL-SIC1*-encoded Sic1 and to reach the *cdc15-2* telophase block. Cells were released at 23°C in YEPD and sampled for immunoblotting and tubulin staining. Clb2-associated kinase activity was measured in separate but identical experiments. **b**, YEPGal (30°C) cultures with the indicated genotypes were resuspended for 30 or 60 min in YEPD to

deplete *GAL-SIC1*-encoded Sic1, chilled and elutriated. The first fraction with greater than 80% budded cells was resuspended into YEPD plus α -factor and sampled for immunoblotting, budding and DAPI-staining. Anti-Sic1 antibody detects a nonspecific band migrating just above Sic1 in *sic1Δ* and *SIC1* samples. Asterisks indicate nonspecific loading control.

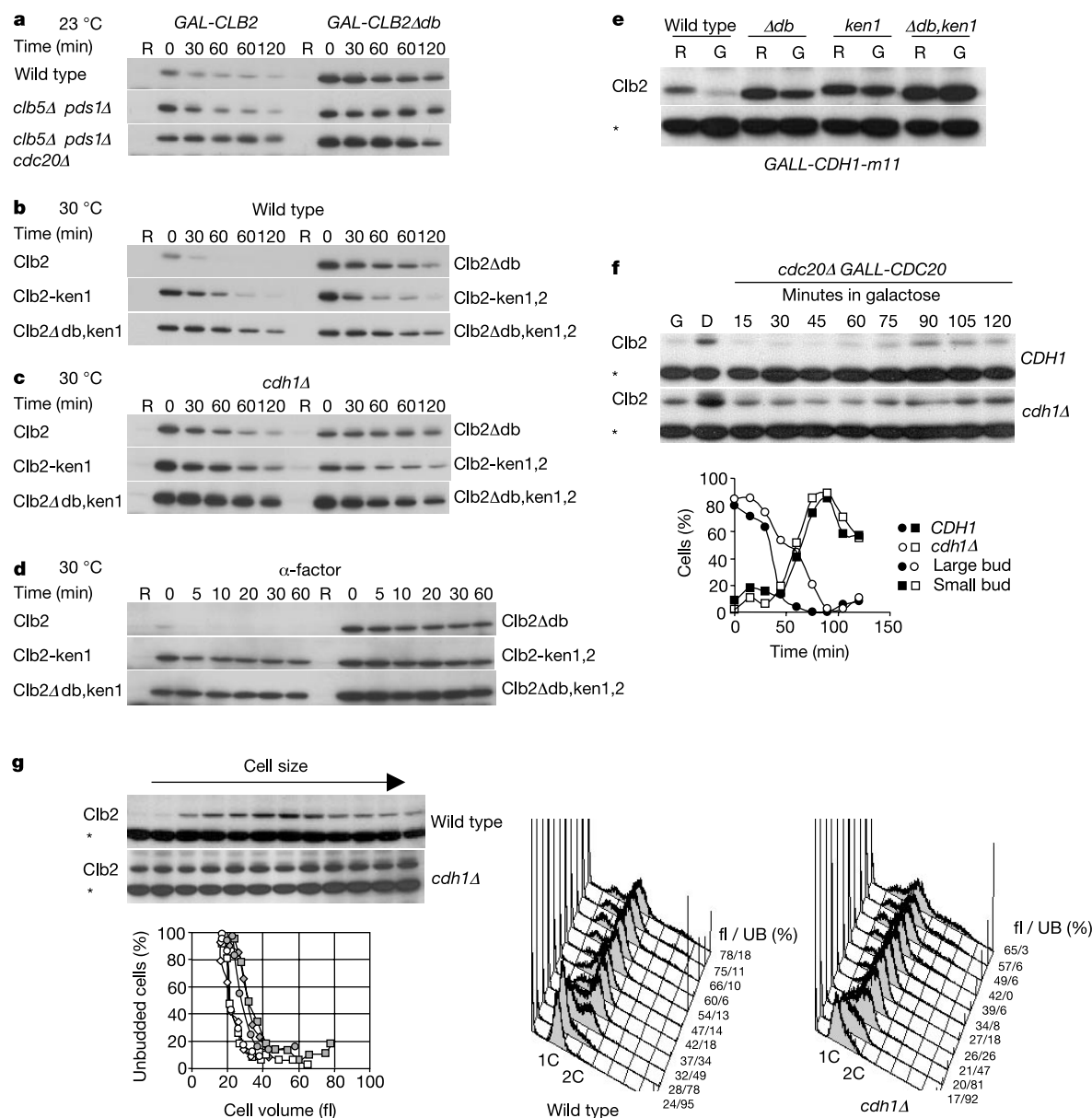


Figure 4 D-box-dependent Clb2 proteolysis depends on Cdc20 and Cdh1, and KEN-dependent proteolysis only on Cdh1. **a–d**, Indicated Clb2 mutants analysed as in Fig. 1a (see Supplementary Information for densitometric quantification). **a**, *CDC20* or *cdc20Δ*; **b**, wild type; **c**, *cdh1Δ*; **d**, α -factor-blocked wild type. **e**, Effect of deregulated Cdh1 (*GALL-CDH1-m11* (ref. 6)); Clb2 mutants are expressed from a truncated *CLB2* promoter

in this context was independent of Cdh1.

KEN boxes, as well as D boxes, control Cdh1-dependent degradation²². Two KEN boxes in Clb2, as well as the D box, were required for efficient Cdh1-dependent degradation (Fig. 4d, e). KEN mutation had minor effects on Clb2 stability in cycling cells (Fig. 4b) and no effect in *cdh1Δ* cells (Fig. 4c), suggesting that Cdc20-dependent Clb2 degradation may be largely KEN-independent. KEN-box mutation in endogenous Clb2 did not prevent viability (data not shown), unlike D-box deletion. D-box deletion may therefore prevent mitotic exit by specifically blocking Cdc20-dependent Clb2 degradation, although an additional effect from interfering with Cdh1-dependent degradation cannot be ruled out.

APC^{Cdc20} may initiate Clb2 degradation²³. Cdc20-dependent Clb2 degradation is not complete, but may be sufficient to allow mitotic exit, even in the absence of Cdh1 and Sic1 (Fig. 5). Sic1 and APC^{Cdh1} provide a second phase of Clb regulation that is not

to allow viability. R, raffinose medium (YEPRaff); G, galactose medium (YEPGal). **f**, Clb2 degradation during mitosis in *cdc20Δ GALL-CDC20* (*CDH1* or *cdh1Δ*) block-release (as in Fig. 1d–f, at 30 °C). **g**, Elutriated wild-type and *cdh1Δ* cultures. Clb2 immunoblots, percentage of unbudded (UB) cells and DNA content in elutriated fractions of the indicated cell volume are shown. Wild type, filled symbols; *cdh1Δ*, open symbols.

required for most aspects of mitotic exit. Although Cdh1/Sic1 regulation may contribute to Clb2–Cdk1 inactivation during mitotic exit, much of the biological significance of this regulation may be in G1, as *cdh1Δ* cells (Fig. 4g) and *sic1Δ* cells¹¹ initiate DNA replication prematurely. The main role of Cdh1 in various eukaryotes may be to maintain a low mitotic cyclin state in G1, perhaps to allow differentiation^{24–26}.

Viability of the *cdc20Δ pds1Δ clb5Δ* strain³ demonstrates the potential for a Cdc20-independent cell cycle oscillator. This oscillator differs from wild type in making the normally dispensable Cdh1 essential³, and it also requires the absence of Pds1 and Clb5, which may antagonize Cdh1 (see Supplementary Information for further discussion). In contrast, the wild-type cell cycle may be primarily governed by a Cdc20-dependent oscillatory mechanism (as in embryonic cell cycles^{24,27,28}), which could involve cyclical Clb-dependent activation of APC^{Cdc20} by phosphorylation^{28,29}.

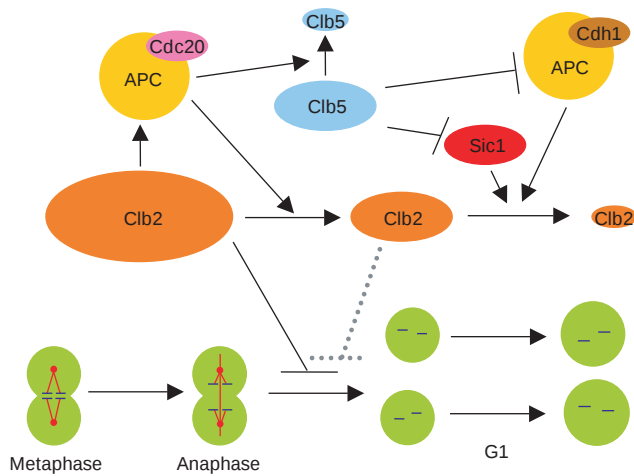


Figure 5 Model of mitotic control. APC^{Cdc20} degrades the mitotic cyclin Clb2 to a level low enough to allow spindle disassembly and cell division. This intermediate level might still interfere with efficient DNA replication in the next cell cycle¹³ (dashed line; see text). APC^{Cdc20}-dependent proteolysis of the S phase cyclin Clb5 prevents Clb5-dependent phosphorylation and inhibition of Cdh1 and the cyclin-dependent kinase inhibitor Sic1 (ref. 3). This allows the phosphatase Cdc14 (not shown) to activate APC^{Cdh1} and Sic1 (ref. 5). APC^{Cdh1} degrades Clb2 further during mitotic exit and in the succeeding G1, and this, combined with inhibition by Sic1, maintains a very low Clb activity level, regulating the length of G1 and probably allowing efficient DNA replication. Cdc14 must have additional targets promoting mitotic exit (see text).

Mutation of Clb2-dependent APC phosphorylation sites²⁹ makes *CDH1*, and the Clb2 KEN boxes recognized by Cdh1, essential (data not shown). Thus, in the absence of APC phosphorylation, a Cdh1-dependent oscillator becomes essential (as in the *cdc20Δ pds1Δ clb5Δ* background³). In the wild-type cell cycle, the Cdh1-dependent oscillator may function predominantly as a mechanism providing G1 stability after exit from mitosis.

Cdc20 and Cdh1 are regulated oppositely by Clb kinase, so these two proposed oscillators may work by entirely different principles. The presence of distinct Cdc20 and Cdh1 orthologues in budding and fission yeast, flies, and humans suggest that this proposed double oscillator may have arisen very early in eukaryotic evolution. □

Methods

Strain and plasmid construction

For details of strain and plasmid construction see Supplementary Information. All strains were of the W303 background. *GALL-CDC20 cdc20Δ pds1Δ* and *clb5Δ::HIS3* were from M. Shirayama. *CLB5* and *CLB2* D-box deletions were described previously^{9,10}. KEN-box mutation of *CLB2* was carried out by polymerase chain reaction mutagenesis and subcloning. The 'ken1' mutation was K100A, E101A, N102A; 'ken1,2' was K85A, E86A, N87A; K100A, E101A, N102A. Replacement of *CLB5* with *CLB5Δdb* was as described for *HA-CLB5Δdb* (ref. 9). Replacement of *CLB2* with *CLB2Δdb* was carried out by integration–excision using *LEU2-CLB2* integrated adjacent to *clb2Δ::URA3*, followed by 5-fluoro-orotic acid selection of pop-outs. Single-copy integration was confirmed by Southern blotting and equal expression levels were compared to a wild-type strain (W303) by northern blotting. The *GAL-SIC1* cassette was from A. Amon. For most experiments a one-copy stable *GAL-SIC1* integrant (1 ×) was used. *cdh1Δ sic1Δ* mutants isolated with this cassette were all diploidized, so for most experiments with *cdh1Δ sic1Δ* strains a two-copy (2 ×) integrant was used. The *GAL-CLB5* and *GAL-CLB2* cassettes were made from the HA-tagged versions¹⁰ by digestion with *NotI* and re-ligation. Therefore, *GAL-CLB* constructs contained three N-terminal amino acids derived from the *NotI* site. Experiments with *CLB5* and *CLB5Δdb* indicate no biological effect of these three amino acids. Cells of the indicated genotype (Fig. 2g) for the plasmid loss assay—also *ade3Δ::URA3* and carrying pLEU2-ADE3* plasmids¹² with one or seven origins of replication—were constructed by tetrad analysis using *ade3Δ::URA3* plasmid-bearing, W303-background strains provided by E. Schwob. All strains also contained a single copy of integrated *GALI::SIC1*. A *clb2Δ::URA3* strain with two integrated copies of *GALL-CDH1-m11* (ref. 6) was transformed with *LEU2*-integrating *CLB2* plasmids. The plasmid contained a truncated *CLB2* promoter, providing low enough expression that *Clb2Δdb* was not lethal. *GALL-CDH1-m11* produces mutant Cdh1 that cannot be inactivated by

phosphorylation, and therefore induces cell cycle arrest and constitutive Clb2 degradation⁶.

Plasmid loss assay

Starter cultures were grown in Gal-Leu medium, then grown in YEPGal for 5 h, and resuspended in YEPD (*GAL-SIC1* off) for 5 h. Before and after YEPD incubation, cultures were streaked out on YEPGal solid medium. The frequency of plasmid loss events was determined as the frequency of white or pink-white, half-sectored (or less than half-sectored) colonies divided by the number of colonies (excluding rare red colonies). The induced frequency of plasmid loss was the frequency of loss after YEPD growth minus the frequency before YEPD growth.

Western blotting and other procedures

Shut-off experiments to measure stability from *GAL-CLB* genes was as described¹⁰, except that *pds1Δ* strains and controls were tested at 23 °C, galactose induction was for 1 h and all strains were resuspended in fresh glucose medium for promoter repression. We used untagged Clb5 because of artefactual effects of tagging on Clb5 stability (Supplementary Information Fig. S2a). In experiments with α-factor blocked cells, cycloheximide (1 mg ml⁻¹) was added at *t* = 0 min. *GALL-CDH1-m11* (ref. 6) strains expressing various *CLB2* mutants (truncated *CLB2* promoter) were grown in raffinose, and Cdh1 was induced by 2.5 h incubation in galactose. DNA content analysis, RNA analysis, elutriation, immunoblotting, H1 kinase assays, and 4,6-diamidino-2-phenylindole (DAPI) and anti-tubulin staining were as described^{10,30}. We established specificity and antibody excess for the Clb2-associated kinase assays by appropriate controls. Anti-Clb antibodies were from Santa Cruz Biotechnology; anti-Sic1 antibody was a gift from M. Tyers. Dilutions were: anti-Clb5 (1:1,000), Clb3 (1:500), Clb2 (1:1,000–1:10,000), Sic1 (1:1,000).

Received 16 October 2001; accepted 3 May 2002; doi:10.1038/nature00856.

- Morgan, D. Regulation of the APC and the exit from mitosis. *Nature Cell Biol.* **1**, E47–E53 (1999).
- Zachariae, W. & Nasmyth, K. Whose end is destruction: cell division and the anaphase-promoting complex. *Genes Dev.* **15**, 2039–2058 (1999).
- Shirayama, M., Toth, A., Galova, M. & Nasmyth, K. APC^{Cdc20} promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. *Nature* **402**, 203–207 (1999).
- Verma, R. *et al.* Phosphorylation of Sic1p by G1 Cdk required for its degradation and entry into S phase. *Science* **278**, 455–460 (1997).
- Visintin, R. *et al.* The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Mol. Cell* **2**, 709–718 (1998).
- Zachariae, W., Schwab, M., Nasmyth, K. & Seufert, W. Control of cyclin ubiquitination by CDK-regulated binding of Hct1 to the anaphase promoting complex. *Science* **282**, 1721–1724 (1998).
- Jaspersen, S., Charles, J. & Morgan, D. Inhibitory phosphorylation of the APC regulator Hct1 is controlled by the kinase Cdc28 and the phosphatase Cdc14. *Curr. Biol.* **9**, 227–236 (1999).
- Lim, H. H., Goh, P. Y. & Surana, U. Cdc20 is essential for the cyclosome-mediated proteolysis of both Pds1 and Clb2 during M phase in budding yeast. *Curr. Biol.* **8**, 231–234 (1998).
- Cross, F. R., Yuste-Rojas, M., Gray, S. & Jacobson, M. D. Specialization and targeting of B-type cyclins. *Mol. Cell* **4**, 11–19 (1999).
- Jacobson, M. D., Gray, S., Yuste-Rojas, M. & Cross, F. R. Testing cyclin specificity in the exit from mitosis. *Mol. Cell Biol.* **20**, 4483–4493 (2000).
- Schwob, E., Böhm, T., Mendenhall, M. D. & Nasmyth, K. The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in *S. cerevisiae*. *Cell* **79**, 233–244 (1994).
- Hogan, E. & Koshland, D. Addition of extra origins of replication to a minichromosome suppresses its mitotic loss in *cdc6* and *cdc14* mutants of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **89**, 3098–3102 (1992).
- Noton, E. & Diffley, J. F. CDK inactivation is the only essential function of the APC/C and the mitotic exit network proteins for origin resetting during mitosis. *Mol. Cell* **5**, 85–95 (2000).
- Schwab, M., Lutum, A. S. & Seufert, W. Yeast Hct1 is a regulator of Clb2 cyclin proteolysis. *Cell* **90**, 683–693 (1997).
- Juang, Y. L. *et al.* APC-mediated proteolysis of Ase1 and the morphogenesis of the mitotic spindle. *Science* **275**, 1311–1314 (1997).
- Visintin, R., Prinz, S. & Amon, A. CDC20 and CDH1: a family of substrate-specific activators of APC-dependent proteolysis. *Science* **278**, 460–463 (1997).
- Huang, J. N., Park, I., Ellingson, E., Littlepage, L. E. & Pellman, D. Activity of the APC(Cdh1) form of the anaphase-promoting complex persists until S phase and prevents the premature expression of Cdc20p. *J. Cell Biol.* **154**, 85–94 (2001).
- Yeong, F. M., Lim, H. H., Wang, Y. & Surana, U. Early expressed Clb proteins allow accumulation of mitotic cyclin by inactivating proteolytic machinery during S phase. *Mol. Cell Biol.* **21**, 5071–5081 (2001).
- Jaspersen, S. L. & Morgan, D. O. Cdc14 activates cdc15 to promote mitotic exit in budding yeast. *Curr. Biol.* **10**, 615–618 (2000).
- Calzada, A., Sacristan, M., Sanchez, E. & Bueno, A. Cdc6 cooperates with Sic1 and Hct1 to inactivate mitotic cyclin-dependent kinases. *Nature* **412**, 355–358 (2001).
- Schwab, M., Neutzner, M., Mockler, D. & Seufert, W. Yeast Hct1 recognizes the mitotic cyclin Clb2 and other substrates of the ubiquitin ligase APC. *EMBO J.* **20**, 5165–5175 (2001).
- Pfleger, C. M. & Kirschner, M. W. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev.* **14**, 655–665 (2000).
- Yeong, F., Lim, H., Padmashree, C. & Surana, U. Exit from mitosis in budding yeast: biphasic inactivation of the Cdc28-Clb2 mitotic kinase and the role of Cdc20. *Mol. Cell* **5**, 501–511 (2000).
- Sigrist, S. J. & Lehner, C. F. *Drosophila* fuzzy-related down-regulates mitotic cyclins and is required for cell proliferation arrest and entry into endocycles. *Cell* **90**, 671–681 (1997).
- Gieffers, C., Peters, B. H., Kramer, E. R., Dotti, C. G. & Peters, J. M. Expression of the CDH1-associated form of the anaphase-promoting complex in postmitotic neurons. *Proc. Natl Acad. Sci. USA* **96**, 11317–11322 (1999).
- Yamaguchi, S., Okayama, H. & Nurse, P. Fission yeast Fizzy-related protein swf1p is a G(1)-specific promoter of mitotic cyclin B degradation. *EMBO J.* **19**, 3968–3977 (2000).

27. King, R. W., Deshaies, R. J., Peters, J.-M. & Kirschner, M. W. How proteolysis drives the cell cycle. *Science* **274**, 1652–1658 (1996).

28. Lorca, T. *et al.* Fizzy is required for activation of the APC/cyclosome in *Xenopus* egg extracts. *EMBO J.* **17**, 3565–3575 (1998).

29. Rudner, A. & Murray, A. Phosphorylation by Cdc28 activates the Cdc20-dependent activity of the anaphase-promoting complex. *J. Cell Biol.* **149**, 1377–1390 (2000).

30. Oehlen, L. J. W. M., McKinney, J. D. & Cross, F. R. Ste12 and Mcm1 regulate cell cycle dependent transcription of FARI. *Mol. Cell. Biol.* **16**, 2830–2837 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank A. Amon, K. Nasmyth, M. Tyers, E. Schwob, W. Seufert and M. Shirayama for reagents and A. Amon, K. Nasmyth, J. Roberts and M. Tyers for useful discussions and critical comments on the manuscript. We also thank C. Li and A. Doty for technical assistance and J. Schmoranz for help with some micrographs. This work was supported by a grant from Deutsche Krebshilfe to R.W. and a PHS grant to F.C.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.C. (e-mail: fcross@rockefeller.edu).

The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair

Karl-Peter Hopfner^{*,†,‡}, Lisa Craig^{*}, Gabriel Moncalian^{*}, Robert A. Zinkel[‡], Takehiko Usui[§], Barbara A. L. Owen^{||}, Annette Karcher[†], Brendan Henderson[¶], Jean-Luc Bodmer^{*}, Cynthia T. McMurray^{||}, James P. Carney[¶], John H. J. Petrini[§] & John A. Tainer^{*,#}

^{*} Department of Molecular Biology and Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA
[†] Gene Center and Institute of Biochemistry, University of Munich, 81377 Munich, Germany
[‡] Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA
[§] Molecular Biology, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA
^{||} Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic and Foundation, Rochester, Minnesota 55905, USA
[¶] The Radiation Oncology Research Laboratory, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA
[#] Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

The Mre11 complex (Mre11–Rad50–Nbs1) is central to chromosomal maintenance and functions in homologous recombination, telomere maintenance and sister chromatid association^{1–7}. These functions all imply that the linked binding of two DNA substrates occurs, although the molecular basis for this process remains unknown. Here we present a 2.2 Å crystal structure of the Rad50 coiled-coil region that reveals an unexpected dimer interface at the apex of the coiled coils in which pairs of conserved Cys–X–X–Cys motifs form interlocking hooks that bind one Zn²⁺ ion. Biochemical, X-ray and electron microscopy data indicate that these hooks can join oppositely protruding Rad50 coiled-coil domains to form a flexible bridge of up to 1,200 Å. This suggests a function for the long insertion in the Rad50 ABC-ATPase domain⁸. The Rad50 hook is functional, because mutations in this motif confer radiation sensitivity in yeast and disrupt binding at the distant Mre11 nuclease interface. These data

support an architectural role for the Rad50 coiled coils in forming metal-mediated bridging complexes between two DNA-binding heads. The resulting assemblies have appropriate lengths and conformational properties to link sister chromatids in homologous recombination and DNA ends in non-homologous end-joining.

Orthologues of the DNA double-strand break repair nuclease Mre11 and the ATPase Rad50 exist in all kingdoms of life and function as a heterotetrameric (Mre11)₂/(Rad50)₂ complex (M2R2)⁸. In eukaryotes, the Mre11 complex can contain a third

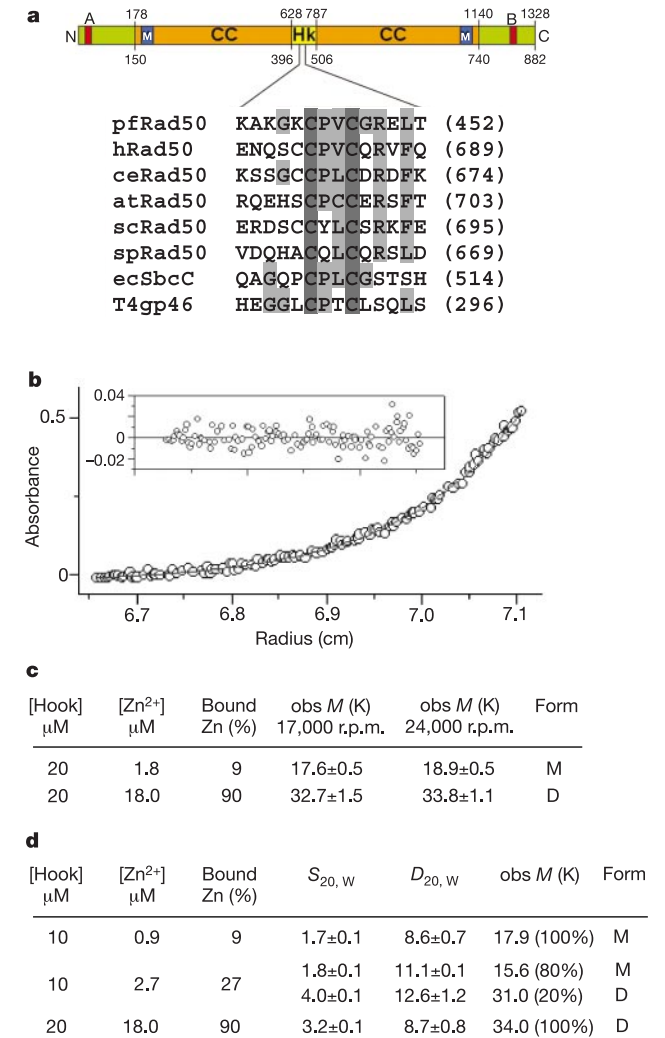


Figure 1 Rad50 domains, sequence conservation of the CXXC motif and Zn²⁺-mediated dimerization of Rad50. **a**, Domain structure of Rad50 and multiple sequence alignment of the central portion of the Rad50 coiled coil showing the conserved CXXC motif. The Walker A (A) and Walker B (B) motifs are labelled in the ABC-ATPase domains (green), and the Mre11-binding site (M) and the hook construct (HK) are labelled in the coiled-coil (CC) regions (orange). Residue numbers delineating the domains and motifs are shown for human (above) and for *P. furiosus* Rad50 (below). Alignment of the region encompassing the CXXC motif is conserved from bacteria to humans (dark grey, invariant cysteine residues; light grey, conserved residues). pf, *P. furiosus*; h, human; ce, *Caenorhabditis elegans*; at, *Arabidopsis thaliana*; sc, *S. cerevisiae*; sp, *Schizosaccharomyces pombe*; ec, *E. coli*; T4, bacteriophage T4. **b**, Representative scan of absorbance versus radius and the residuals (inset) for the equilibrium analysis of pfRad50–CXXC–L in the presence of Zn²⁺. **c**, **d**, Sedimentation equilibrium analysis (**c**) and sedimentation velocity analysis (**d**) of pfRad50–CXXC–L in the presence or absence of Zn²⁺. Zn²⁺ concentration was determined by atomic absorption; obs *M* is the observed *M* (K) at 17,000 or 24,000 r.p.m.; *S*_{20,w} is the sedimentation coefficient (10^{–13} s); *D*_{20,w} is the diffusion constant (cm² s^{–1}). The theoretical *M* deduced from the sequence is 16,628 for the monomer (M) and 33,256 for the dimer (D).

component, named Xrs2 in yeast and Nbs1 in vertebrates⁹. The Mre11 complex is involved in meiotic DNA processing, homologous recombination, telomere maintenance, non-homologous end-joining and DNA damage detection^{1,10}. Purified Mre11 complex degrades DNA ends and DNA hairpins in an ATP-modulated manner, an activity that is required to process misfolded DNA at broken ends^{11–16}. But mutations that abrogate this nuclease activity do not affect recombination and telomere maintenance^{2,3,5}. By contrast, null mutations of Mre11, Rad50 and Nbs1 are nonviable in vertebrates and show severe recombination and telomere maintenance defects in yeast^{4,17–19}. These observations indicate that the Mre11 complex has an important function that is distinct from its DNA end-processing activity. Indeed, studies have suggested that the Mre11 complex acts in tethering DNA ends^{20,21}, in the proper formation of synaptonemal complexes⁶ and in sister chromatid interactions⁷. But the molecular mechanisms by which the Mre11 complex joins DNA segments and ends have not been established.

The most striking feature of Rad50 in all species is a coiled-coil sequence of 150–600 Å that links the two parts of the folded ABC-ATPase domain (ref. 22 and Fig. 1a). At one end of the bipolar M2R2 complex, the coiled coils are attached to the Rad50 ATPase domains and contain the interaction sites for Mre11 (refs 8, 20, 21, 23, 24 and Fig. 1a). In the centre of the coiled-coil sequence is a conserved Cys-X-X-Cys (CXXC) motif that has been proposed to form a protein

interaction motif (ref. 25 and Fig. 1a). To examine the role of the CXXC motif in the function of the Mre11 complex, we expressed and characterized coiled-coil polypeptides encompassing this region in *Pyrococcus furiosus* (pf), *Saccharomyces cerevisiae* (sc) and human (h) Rad50 (denoted Rad50-CXXC).

To test whether Rad50-CXXC could form dimers through coordination of a Zn^{2+} ion by the four conserved cysteines, we carried out gel filtration and analytical ultracentrifugation on pfRad50-CXXC-L (residues 381–522, calculated molecular mass 16,600; M 16.6K). pfRad50-CXXC-L elutes as two fractions in gel-filtration chromatography with an apparent M of 17K and 35K, which is consistent with monomeric and dimeric forms (data not shown). Protein eluting in the dimer fraction could be eluted in the monomer fraction after treatment with chelating and reducing agents. Analytical ultracentrifugation precisely determined that Rad50-CXXC was monomeric ($M \approx 18K$) in the absence of Zn^{2+} , and dimeric ($M \approx 33K$) on addition of Zn^{2+} (Fig. 1b, c). In addition, the 127-residue scRad50-CXXC polypeptide co-immunoprecipitated with full-length scRad50 when co-expressed in yeast. Similarly, the hRad50-CXXC dimer contained Zn^{2+} in a molar ratio of 0.4 Zn^{2+} atoms per hRad50-CXXC molecule, as determined by atomic absorption spectroscopy, and Zn^{2+} was not removed by exhaustive dialysis against chelating agents (data not shown). Together, these results suggest that the CXXC motif promotes a

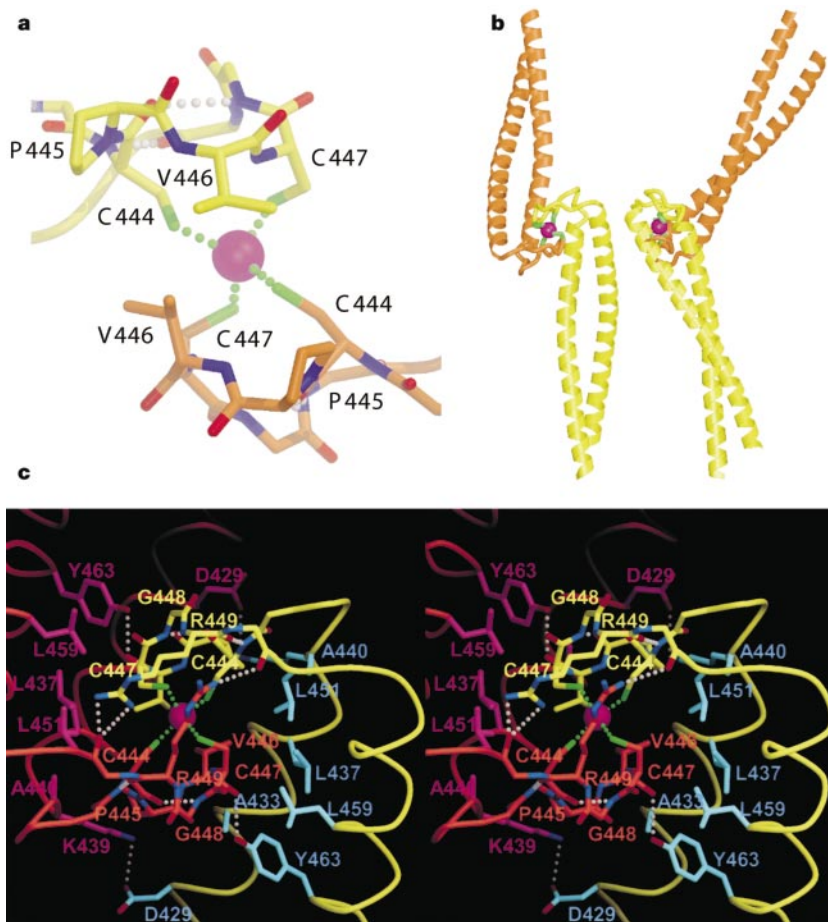


Figure 2 Crystal structures of the central portion of the Rad50 coiled coil reveal a dimerization motif mediated by a single Zn^{2+} -binding site. **a**, Two pfRad50-CXXC molecules (yellow and orange) form a composite Zn^{2+} -binding site at the dimer interface. Each pfRad50-CXXC contributes two cysteines (green) to tetrahedrally coordinate the metal (Zn^{2+} or Hg^{2+} , magenta sphere). The interaction is stabilized by two main-chain hydrogen bonds (white dots). **b**, Fold and assembly of the metal-bound pfRad50-CXXC dimer. Each of the two pfRad50-CXXC molecules forms an antiparallel coiled coil around a 14-residue non-helical hook-shaped region, which encompasses the CXXC motif and

folds into the Zn^{2+} -binding hook. The coiled coils extend from the composite Zn^{2+} -binding site in opposite directions, which allows them to link two distinct Mre11 complexes. The dimer on the right is rotated 90° relative to the one on the left. **c**, Stereo view of the Rad50-CXXC dimer interface showing the two cysteine hooks, bound metal ion and hydrophobic interactions. Hydrophobic side chains are labelled and coloured pink for one pfRad50-CXXC and blue for the other. Additional hydrogen bonds and salt bridges are indicated by white dots.

metal-dependent interaction between Rad50 molecules *in vivo* and *in vitro*.

To reveal the mechanistic role of the Rad50-CXXC motif and its metal-dependent dimerization, we crystallized pfRad50-CXXC (residues 396–506) in the presence of Zn^{2+} and Hg^{2+} and solved their crystal structures to 2.6 and 2.2 Å, respectively. The structure shows that the CXXC motif forms a molecular hook assembly at the apex of the Rad50 coiled coil. This two-cysteine hook creates one-half of a composite Zn^{2+} -binding site that links two Rad50-CXXC molecules (Fig. 2a). Residues 396–439 and 454–498 each form 13 helical turns, resulting in an antiparallel coiled coil that is 72 Å long, 18 Å wide and 12 Å thick (Fig. 2b). The central hook structure consists of 14 non-helical residues (440–453) with two main-chain hydrogen bonds. This hook structure forms an antiparallel extension at the coiled-coil apex that protrudes 15 Å perpendicular to the flat side of the coiled coil (Fig. 2b).

Each hook motif of the two Rad50-CXXC molecules contributes two cysteine residues (Cys 444 and Cys 447) to the metal-binding site that together coordinate tetrahedrally the Zn^{2+} (or Hg^{2+}) ion (Fig. 2a). The metal site contributes $\sim 70 \text{ Å}^2$ of the $\sim 1,200 \text{ Å}^2$ total surface area buried by the two interlocking hook extensions of the Rad50-CXXC dimer. The dimer assembly is stabilized by a small conserved hydrophobic interface: Pro 445 and Val 446 of the CXXC motif bind in *trans* to a hydrophobic groove formed by Ala 433, Leu 437, Leu 451 and Leu 459 (Fig. 2c). This hydrophobic groove lies in the cleft between the coiled coil and the hook extension. The conserved hydrophobic Leu 451 (leucine or phenylalanine in most Rad50 sequences) at the beginning of the hook extension points into the centre of the hydrophobic interface. This central location suggests that Leu 451 stabilizes the perpendicular structure of the hook extension that caps the coiled coil. The dimer assembly is stabilized further by intermolecular hydrogen bonds, including those from the side chains of Arg 449 and Tyr 463 to the main-chain carbonyl of Val 446, and a salt bridge between the side chains of Lys 439 and Asp 429.

The structure of the Rad50-CXXC dimer has consequences for

the structure and biological function of the Mre11 complex. First, the intramolecular coiled coil observed in the crystal structures would produce a Rad50 ATPase domain assembled from the amino- and carboxy-terminal ATPase segments of a single Rad50 polypeptide chain. Second, the two coiled-coil regions extend from the Zn^{2+} -binding site in almost opposite directions ($\sim 140^\circ$ angle, Fig. 2b). This arrangement indicates that the hook structure can link coiled coils of two different, oppositely protruding Mre11 complexes, which results in a dimer of Mre11 complex heterotetramers with the stoichiometry $(M2R2)_2$ (Fig. 3l).

To investigate the interactions mediated by the cysteine hook between Mre11 complexes, we imaged human and *P. furiosus* Mre11 complexes in the presence and absence of Zn^{2+} by electron microscopy (Fig. 3). In the absence of Zn^{2+} , we observed asymmetric, lollipop-shaped Mre11 complexes with dimensions consistent with heterotetrameric M2R2 stoichiometry, as reported previously (ref. 8 and Fig. 3a, b). Notably, in some of the M2R2 complexes the two Rad50 coiled-coil arms were separated, showing a coiled coil formed by a single polypeptide chain (Fig. 3c, d) consistent with the X-ray structures of Rad50-CXXC. This arrangement precludes the assembly of ATPase domains through swapped, intermolecular coiled coils from two Rad50 molecules, as we had proposed previously^{8,22} and as also proposed for the scRad50 complex²⁴.

In the presence of Zn^{2+} , we observed several different Mre11 complex forms that were all consistent with Zn^{2+} -mediated linkage of the Rad50 hooks. As predicted from the metal-linked Rad50-CXXC crystal structures, we observed dumbbell-shaped particles with dimensions consistent with two M2R2 heterotetramers linked tail-to-tail through their cysteine hook motifs (Fig. 3e–g). As the coiled coils of these octameric complexes are flexible, existing in both extended and highly curved conformations, they could potentially bring two DNA ends within direct proximity of each other. Occasionally, we observed two M2R2 complexes joined by both a tail-to-tail linkage and a head-to-head interaction (Fig. 3h). Such a

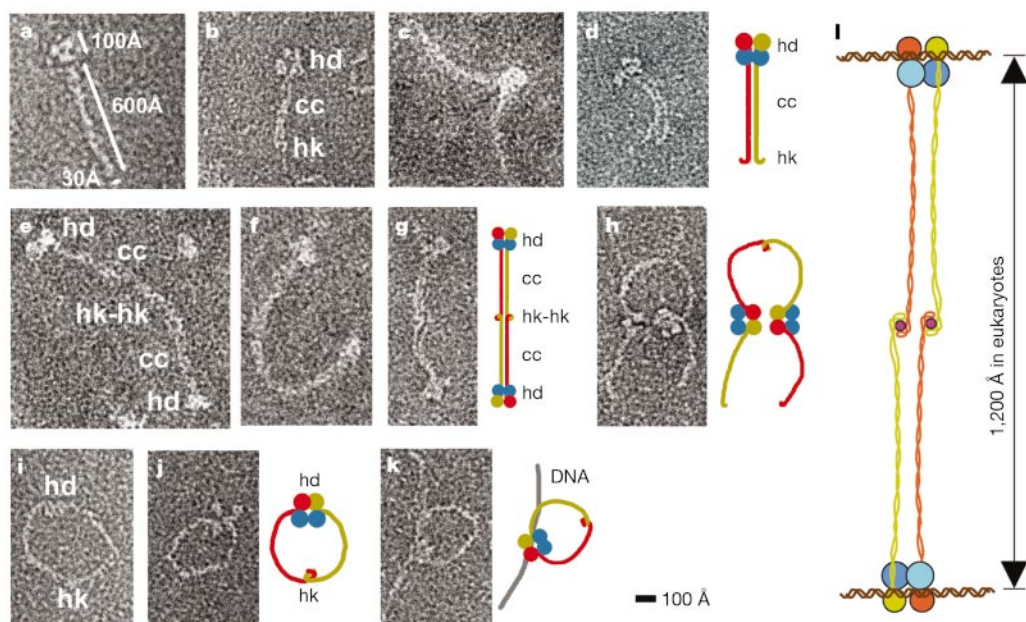


Figure 3 Negatively stained electron micrographs of human and *P. furiosus* Mre11 complexes reveal conformational properties of the component domains including Zn^{2+} -mediated tail-to-tail linkages between individual coiled-coil arms of Rad50. **a, c**, Human M2R2 tetramer; **b, d**, *P. furiosus* M2R2 tetramer; **e, f**, hook-linked human $(M2R2)_2$ octamer; **g**, hook-linked *P. furiosus* $(M2R2)_2$ octamer; **h**, single-linkage $(M2R2)_2$ octamer; **i**, circular human M2R2 tetramer; **j**, circular *P. furiosus* M2R2 tetramer; **k**, circular *P. furiosus* M2R2 tetramer bound to DNA. hd, head domain comprising two Rad50 ATPases (red and yellow

spheres in the diagrams) and two Mre11 molecules (blue spheres); cc, coiled coil (single red or yellow line); hk, Zn^{2+} -hook. **l**, Architecture of the Zn^{2+} -linked $(M2R2)_2$ Mre11 complex and its interaction with DNA. An Mre11 dimer (blue spheres) binds to the coiled coils of two Rad50 molecules adjacent to the ATPase domains (red/yellow spheres), forming the DNA-binding head of the Mre11 complex. The two binding heads are joined by a Zn^{2+} -mediated tail-to-tail linkage between the coiled-coil hooks of Rad50, providing a bridge between sister chromatids before recombination. Zn^{2+} is shown as a magenta sphere.

linkage could potentially form a metal-linked zipper or network of Mre11 complexes between two linked DNA segments, as is suggested by the high flexibility of the coiled-coil arms and the tendency of the Mre11 complex to form clusters at DNA ends²⁰.

Besides the hook-linked (M2R2)₂ octamers, we occasionally observed individual Mre11 complexes that were circular, with a linkage formed between the two Rad50 coiled coils of the same complex (Fig. 3i–k). Similar particles have been observed in scanning atomic force microscopy of human Mre11 complex²⁰ and in rotary shadowed electron microscopy of scRad50 and Mre11 complex²⁴. The architecture of the hook-linked (M2R2)₂ octamer, on the basis of the electron microscopy and crystallographic data, is appropriate for bridging two distinct M2R2 heads. These results therefore suggest the molecular basis for the role of Rad50 in sister chromatid interactions during homologous recombination⁷ and in proper formation of synaptonemal complexes (ref. 6 and Fig. 3l).

The hook-linked (M2R2)₂ octamers and circular M2R2 tetramers were rarely seen in the absence of Zn²⁺. They were abundant in the presence of exogenous Zn²⁺ (with 40% of human Mre11 complexes showing the (M2R2)₂ form in some preparations), however, which suggests that their interaction is metal-mediated rather than a folding artefact. Notably, the hook-linked forms were observed

much more frequently in hRad50 protein than in pRad50, which suggests that the recombinant pRad50 that was overexpressed in *Escherichia coli* has a lower affinity for Zn²⁺; this might reflect a metal-rich environment of the *Pyrococcus* or differences in the expression and purification methods.

To assess the functional significance of the CXXC motif in recombinational DNA repair, we mutated individually the two conserved cysteine residues in the scRad50 C₁CXXC₂ motif to glycine, and analysed the stability of these mutants in terms of their plating efficiency and their ability to rescue the ionizing radiation sensitivity phenotype of a Rad50-deficient strain. The plating efficiency of both the C₁G and C₂G mutants was reduced by a factor of 3–8 relative to the wild-type transformants and was essentially equivalent to that of the Rad50-deficient strain. Ionizing radiation sensitivity of both mutants was increased relative to the wild-type strain, with the C₁G mutant showing a fivefold increase and the C₂G mutant showing a 95-fold increase at 300 Gy (Fig. 4a).

Because mutations that disrupt complex assembly show a null Mre11 complex phenotype, we examined the ability of the C₁G and C₂G Rad50 mutants to bind Mre11. The Mre11–Rad50 interaction was disrupted in the C₂G mutant, which suggests a defect in the formation of the Mre11-binding site at the structurally distinct base of the Rad50 coiled coil (Fig. 4a). By contrast, the C₁G mutant did not impair Rad50–Mre11 interaction, as indicated by the presence of mutant Rad50 in Mre11 immunoprecipitates, which suggests that the increased radiation sensitivity in this mutant is a direct effect of altering the cysteine hook Zn²⁺-binding site. Introducing glycine into the CXXC motif in the C₁G mutant would increase backbone flexibility of the hook region, which might allow the additional cysteine adjacent to C₁ (C₁CXXC₂) in scRad50 to rescue partially Zn²⁺ binding *in vivo*. This region is presumably more flexible in scRad50 because it lacks the proline that is present in many Rad50 hooks (Fig. 1a). This partial rescue may account for the only fivefold increase in ionizing radiation sensitivity observed in the C₁G mutant. In support of this, a C₁S mutant, which would have less backbone flexibility than C₁G, disrupted the Mre11 complex and showed ionizing radiation sensitivity equivalent to that of the C₂G mutant. A C₂S mutant, like the C₂G, also disrupted Mre11 interaction (data not shown). Taken together, these data indicate that the CXXC motif is important in Mre11 complex assembly as well as in recombinational DNA repair.

How might the Mre11 complex integrate binding events involving two DNA ends? In principle, some form of coupling between the Mre11 complex DNA-binding heads is required. Thus, the property of the Rad50 hook motifs to form such a linkage provides a molecular basis for the essential, nuclease-independent function of the Mre11 complex in DNA recombination and repair. The experimentally defined distance of 1,200 Å between the head domains of two Zn²⁺-linked human M2R2 complexes (Fig. 3e,f) is appropriate to bridge two sister chromatids before recombination (Fig. 4b) and/or to bridge two separated DNA ends after occurrence of a DNA double-strand break (Fig. 4b, c). Thus, a structurally implied primary activity of the Mre11 complex seems to be a linker function for homologous stretches of DNA and/or broken ends. For non-homologous end-joining, two broken ends might be brought together by the collapse of a Zn²⁺-linked (M2R2)₂ octamer (Fig. 3f, h and Fig. 4c). Alternatively, two broken ends might bind to the head of the same Mre11 complex that has an internal Zn²⁺ linkage (Fig. 3i–k and Fig. 4d). The Rad50 hook conformation is evidently linked to the Rad50 DNA-binding heads, which undergo ATP-driven rotations²². These results thus suggest how the Rad50 hook may integrate DNA binding by two ostensibly independent Mre11 complexes. Consequently, the hook-linked Mre11 complexes identified here provide a specific and testable molecular mechanism for Rad50 to function as an ATP-modulated DNA crosslinker, as has been proposed for the related but distinct SMC proteins²⁶. □

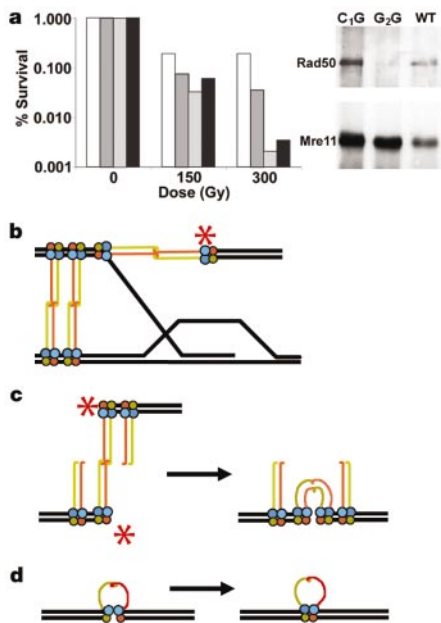


Figure 4 *In vivo* effects of cysteine mutations in the CXXC motif and biological implications for metal-mediated dimerization of Rad50. **a**, Ionizing radiation sensitivity (left) and Mre11 interactions (right) of Rad50 transformants in a Rad50-deficient yeast strain. Left, cells were γ -irradiated in suspension at 150 or 300 Gy and plated in triplicate. Wild type, white; C₁G mutant, dark grey; C₂G mutant, light grey; empty vector, black. Survival relative to the corresponding non-irradiated strain, corrected for plating efficiency, is plotted. Individual data are given in Supplementary Information Table 2. Right, immunoprecipitation of Rad50 mutants by Mre11 antiserum from cell extracts of the indicated transformants. Immunoprecipitates were fractionated by SDS-PAGE and immunoblotted with antisera against Rad50 (top) or Mre11 (bottom). **b–d**, Models for the role of the Mre11 complex in homologous recombination and non-homologous end-joining. **b**, The dsDNA nuclease activity of the Mre11 complex may process DNA ends for subsequent repair steps (asterisk). The architecture of the tail-to-tail linked Mre11 complexes suggests a mechanism for bridging sister chromatids, and initiating and stabilizing displacement loop formation—a common intermediate in recombination, break-induced replication and telomere maintenance. **c**, In non-homologous end-joining, the head domains of two M2R2 complexes bind separate broken DNA ends, aligning and tethering them after a structural transition. This conformational flexibility matches the different shapes of the M2R2 complex in electron microscopy. Asterisk indicates potential nucleolytic activity of the M2R2 complex. **d**, As observed in electron microscopy, circularization of a single M2R2 complex could potentially contribute to productive repair of broken DNA ends.

Methods

Protein expression and purification

P. furiosus Rad50-CXXC (residues 396–506, MKLLLEETKKTITIEERNEITQRIGELK NKIGDLKTAIEELKKAAGKCPVCGRELTDEHRELLSKYHLDLNNKNTLAKLIDRK SELERELRRIDMEIKRLTPLTVAE) was expressed in *E. coli* BL21-RIL (DE3) (Stratagene) using the pET29 vector (Novagen) and purified to homogeneity by heat denaturation, HiTrap SP (Pharmacia) cation exchange chromatography and Superdex S75 (Pharmacia) gel-filtration chromatography. A longer polypeptide, pRad50-CXXC-L (residues 381–522, RQLKEKLGDKSPEDIKKLLEETKKTITIEERNEITQRIGELKKNIGDLKTAIEEL KKAAGKCPVCGRELTDEHRELLSKYHLDLNNKNTLAKLIDRKSELERELRRIDMEI KRLTPLTVAEQIRSIIEELNVVLEK) was used for gel-filtration and analytical ultracentrifugation studies. Expression of the human and *P. furiosus* Mre11 complexes has been described⁸.

Analytical ultracentrifugation

Protein from the gel-filtration peak corresponding to the pRad50-CXXC-L dimer was exchanged into chelating (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM dithiothreitol (DTT), 150 μ M EDTA) or Zn^{2+} -containing (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 150 μ M zinc acetate, 500 μ M Tris-(2-carboxyethyl) phosphine hydrochloride) buffer, respectively, using Bio-Spin 6 columns at room temperature. We determined bound Zn^{2+} by atomic absorption. Sedimentation equilibrium measurements were made on an Optima XL-I equipped with an ultraviolet/interference detection system (Beckman Instruments) as described²⁷. Experiments were carried out at 20 °C in an ANTi60 rotor until equilibrium was achieved. We analysed each sample at 17,000, 24,000 and 35,000 r.p.m. to optimize the detection of tetrameric, dimeric and monomeric species, respectively. Fitting was carried out by a nonlinear Levenberg–Marquardt routine.

Crystallography

pRad50-CXXC (30 mg ml⁻¹ in 20 mM phosphate (pH 7.5), 200 mM NaCl, 0.1 mM EDTA, 5% glycerol, 2 mM DTT) was crystallized in space group *P*₂ with cell dimensions *a* = 32.0, *b* = 78.0, *c* = 53.3 Å, β = 91.6° and two molecules in the asymmetric unit, by sitting-drop vapour diffusion after mixing 2 μ l of protein solution with 1.5 μ l of precipitant solution (100 mM phosphate/citrate pH 4.2, 40% ethanol, 10% glycerol, 20 mM zinc acetate). Specific binding of mercury acetate to Rad50-CXXC under crystallization conditions was verified by native PAGE and used for co-crystallization.

For Rad50-CXXC co-crystallized with 1 mM mercury acetate, two-wavelength multiple anomalous diffraction data to 2.2 Å were recorded at 100 K with a Quantum 2 × 2 CCD detector at BL 9-2 (SSRL) and processed with HKL (Z. Otwinowski; Supplementary Information Table 1). A single peak was located in the anomalous difference Patterson map and used to calculate phases with MLPHARE to 3 Å resolution. After 15 cycles of solvent flipping with SOLOMON, the resulting high-quality electron density was used to trace both molecules in the asymmetric unit with MAIN (D. Turk). After overall anisotropic *B*-value and bulk solvent correction, the structure was refined to 2.2 Å by cycles of positional and individual *B*-value refinement and manual model building with CNS²⁸ and MAIN (Supplementary Information Table 1). Five per cent of the data were excluded from refinement to calculate the free *R* value for crossvalidation. Non-crystallographic symmetry (NCS) restraints were gradually removed in later stages of refinement. The current model is refined to an *R* factor of 0.225 (*R*_{free} 0.278) at 2.2 Å resolution using 10,580 unique reflections (88% completeness). Refinement and model statistics are given in Supplementary Information Table 1. The refined coordinates of Hg²⁺-Rad50-CXXC were used to solve the 2.6 Å structure of Zn^{2+} -Rad50-CXXC by molecular replacement. Both structures are virtually identical, including the metal coordination site, so the higher quality Hg²⁺-Rad50-CXXC is shown here.

Electron microscopy

Carbon-coated copper grids were glow-discharged in the presence of amylamine and floated for 2 min on 5- μ l drops containing 50 μ g ml⁻¹ protein pre-incubated with either 0.5 mM zinc acetate or buffer lacking Zn^{2+} . We stained the grids with 3% uranyl acetate and viewed them on a Philips CM100 microscope at 100 kV in low-dose mode. In some cases (Fig. 3c, d), 0.5% trehalose was added to the sample before staining to increase the contrast.

Mutational analysis in yeast

The RAD50 open reading frame (ORF), as well as the C₁G and C₂G rad50 mutants, were expressed from the PHO5 promoter in YEplac195. Mutations were introduced as described⁹ and confirmed by sequencing the whole ORE. We grew transformants in standard phosphate conditions to limit overexpression from the PHO5 promoter. Ionizing radiation sensitivity was assessed as described²⁹. We prepared protein extracts and carried out immunoprecipitation with Mre11 antiserum as described³⁰. Western blotting confirmed that mutant Rad50 proteins and Mre11 were present in normal amounts, irrespective of complex formation (data not shown).

Received 18 April; accepted 28 May 2002; doi:10.1038/nature00922.

- Haber, J. E. The many interfaces of Mre11. *Cell* **95**, 583–586 (1998).
- Bressan, D. A., Baxter, B. K. & Petrini, J. H. The Mre11–Rad50–Xrs2 protein complex facilitates homologous recombination-based double-strand break repair in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 7681–7687 (1999).
- Moreau, S., Ferguson, J. R. & Symington, L. S. The nuclease activity of Mre11 is required for meiosis but not for mating type switching, end joining, or telomere maintenance. *Mol. Cell. Biol.* **19**, 556–566 (1999).
- Yamaguchi-Iwai, Y. *et al.* Mre11 is essential for the maintenance of chromosomal DNA in vertebrate cells. *EMBO J.* **18**, 6619–6629 (1999).

- Tsukamoto, Y., Taggart, A. K. & Zakian, V. A. The role of the Mre11–Rad50–Xrs2 complex in telomerase-mediated lengthening of *Saccharomyces cerevisiae* telomeres. *Curr. Biol.* **11**, 1328–1335 (2001).
- Merino, S. T., Cummings, W. J., Acharya, S. N. & Zolan, M. E. Replication-dependent early meiotic requirement for Spo11 and Rad50. *Proc. Natl Acad. Sci. USA* **97**, 10477–10482 (2000).
- Hartsuiker, E., Vaessen, E., Carr, A. M. & Kohli, J. Fission yeast Rad50 stimulates sister chromatid recombination and links cohesion with repair. *EMBO J.* **20**, 6660–6671 (2001).
- Hopfner, K. P. *et al.* Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105**, 473–485 (2001).
- Petrini, J. H. The Mre11 complex and ATM: collaborating to navigate S phase. *Curr. Opin. Cell Biol.* **12**, 293–296 (2000).
- Huang, J. & Dynan, W. S. Reconstitution of the mammalian DNA double-strand break end-joining reaction reveals a requirement for an Mre11/Rad50/NBS1-containing fraction. *Nucleic Acids Res.* **30**, 667–674 (2002).
- Furuse, M. *et al.* Distinct roles of two separable *in vitro* activities of yeast Mre11 in mitotic and meiotic recombination. *Embo J* **17**, 6412–6425 (1998).
- Paull, T. T. & Gellert, M. The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Mol. Cell* **1**, 969–979 (1998).
- Connelly, J. C., de Leau, E. S. & Leach, D. R. DNA cleavage and degradation by the SbcCD protein complex from *Escherichia coli*. *Nucleic Acids Res.* **27**, 1039–1046 (1999).
- Hopfner, K. P. *et al.* Mre11 and Rad50 from *Pyrococcus furiosus*: cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine. *J. Bacteriol.* **182**, 6036–6041 (2000).
- Trujillo, K. M. & Sung, P. DNA structure-specific nuclease activities in the *Saccharomyces cerevisiae* Rad50/Mre11 complex. *J. Biol. Chem.* **276**, 35458–35464 (2001).
- Lobachev, K. S., Gordonin, D. A. & Resnick, M. A. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **108**, 183–193 (2002).
- Xiao, Y. & Weaver, D. T. Conditional gene targeted deletion by Cre recombinase demonstrates the requirement for the double-strand break repair Mre11 protein in murine embryonic stem cells. *Nucleic Acids Res.* **25**, 2985–2991 (1997).
- Luo, G. *et al.* Disruption of mRad50 causes embryonic stem cell lethality, abnormal embryonic development, and sensitivity to ionizing radiation. *Proc. Natl Acad. Sci. USA* **96**, 7376–7381 (1999).
- Zhu, J., Petersen, S., Tessarollo, L. & Nussenzweig, A. Targeted disruption of the Nijmegen breakage syndrome gene NBS1 leads to early embryonic lethality in mice. *Curr. Biol.* **11**, 105–109 (2001).
- de Jager, M. *et al.* Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol. Cell* **8**, 1129–1135 (2001).
- Chen, L., Trujillo, K., Ramos, W., Sung, P. & Tomkinson, A. E. Promotion of DnI4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. *Mol. Cell* **8**, 1105–1115 (2001).
- Hopfner, K. P. *et al.* Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101**, 789–800 (2000).
- Connelly, J. C., Kirkham, L. A. & Leach, D. R. The SbcCD nuclease of *Escherichia coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc. Natl Acad. Sci. USA* **95**, 7969–7974 (1998).
- Anderson, D. E., Trujillo, K. M., Sung, P. & Erickson, H. P. Structure of the Rad50 × Mre11 DNA repair complex from *Saccharomyces cerevisiae* by electron microscopy. *J. Biol. Chem.* **276**, 37027–37033 (2001).
- Sharples, G. J. & Leach, D. R. Structural and functional similarities between the SbcCD proteins of *Escherichia coli* and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. *Mol. Microbiol.* **17**, 1215–1217 (1995).
- Hirano, T. SMC-mediated chromosome mechanics: a conserved scheme from bacteria to vertebrates? *Genes Dev.* **13**, 11–19 (1999).
- Wu, X. & McMurray, C. T. Calmodulin kinase II attenuation of gene transcription by preventing cAMP response element-binding protein (CREB) dimerization and binding of the CREB-binding protein. *J. Biol. Chem.* **276**, 1735–1741 (2001).
- Brunker, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Bressan, D. A., Olivares, H. A., Nelms, B. E. & Petrini, J. H. Alteration of N-terminal phosphoesterase signature motifs inactivates *Saccharomyces cerevisiae* Mre11. *Genetics* **150**, 591–600 (1998).
- Usui, T., Ogawa, H. & Petrini, J. H. A DNA damage response pathway controlled by Tel1 and the Mre11 complex. *Mol. Cell* **7**, 1255–1266 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank D. Turk for MAIN; M. Gehl for atomic absorption spectroscopy analysis and A.-M. Hayes for technical advice. We acknowledge the key facilities and staff of synchrotron beamlines BL9-2 (SSRL) and 5.0.2 (ALS). Key funding was provided by the US Department of Energy (DOE) and NCI (J.A.T. and J.P.C.), the American Cancer Society (J.P.C.), the Canadian Institutes of Health Research (L.C.), National Institutes of Health, Human Frontiers Science Program and DOE (J.H.J.P.), Skaggs Institute for Chemical Biology (K.P.H.), the National Institutes of Mental Health (C.T.M.) and The Swiss National Science Foundation (J.L.B.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.T. (e-mail: jat@scripps.edu) or J.H.J.P. (e-mail: petrini@mskcc.org). Coordinates for pRad50-CXXC have been deposited at the Protein Data Bank under accession code 1L8D.

The voyage of discovery

Data mining, susceptibility testing, high-throughput screening for drugs.

ProtoZONE reader

Synbiosis www.synbiosis.com

Testing times

The ProtoZONE system is a dedicated inhibition zone reader that automatically measures zones to produce antibiotic susceptibility test results in seconds. The integrated system includes PC technology and a CCD camera to analyse sample plates containing antibiotic disks and their related inhibition zones. The software sizes the zones around each disk and a high-resolution image is displayed for each plate, with susceptibility flags shown next to the measured zones. Synbiosis says that ProtoZONE can read, display and store zone measurements for a 9-cm plate with six disks in under 2 seconds. The software can be used with most agar types, including zones on blood agar that are difficult to see.

AquaMax

Molecular Devices www.moleculardevices.com

You can dispense with this

The AquaMax 96/384-well liquid-handling system was developed to provide a quick, simple and reliable method of dispensing solutions, beads and cells prior to assay and analysis. It can dispense volumes as low as 0.5 µl into 384-well microplates. Throughput is rapid at less than 30 seconds per plate.

Multi-Gen 7

PRO Scientific www.proscientific.com

Steel yourself

Multi-Gen 7 is the name of PRO Scientific's series of rotor stator generators designed to homogenize many samples without the risk of cross-contamination. Available in packages of 24, the 7-mm generators attach to the motor unit via a motor collar that allows hands-free ejection of the used generator.



Get a grip: mixing with Multi-Gen 7.



Screen test: ProtoZONE gives a visual check on antibiotic susceptibility.

Multi-Gen 7 generators are made of stainless steel, allowing repeated autoclaving and homogenizing of tough tissue.

eLabNotebook

Beckman Coulter www.beckmancoulter.com

Web-based application resource

eLabNotebook is an applications-based service recently added to Beckman Coulter's web site. It offers information on specific applications for the company's life science research products, focusing on drug discovery. All materials for an application are located together on the web site, including methods, application bulletins, product specifications, troubleshooting information and FAQs.

HT-12 solvent evaporator

Genevac www.genevac.com

It's an automatic turn-off

Genevac's enhanced HT-12 large-scale solvent evaporator features new control software that allows the instrument to automatically detect when samples are dry, and shut itself down. Combined with the company's VC3000TA continuous running condensers, this allows unattended operation and overnight evaporation runs. Large numbers of samples of broadly similar mass and format, such as microplates and HPLC autosampler vials, can be dried together since the triple-deck rotor can accommodate 48 shallow-well plates simultaneously. The

evaporator is compact enough to be used in any standard fume hood or on a dedicated trolley.

CellSpace 2.0

Cellomics www.cellomics.com

All mine

The CellSpace Version 2.0 knowledge mining system continuously monitors 642 journals covering molecular and cellular biology, clinical chemistry and pathology. It has a database of over 1.8 million papers. Users can determine the biological functions most commonly shared by a cluster of gene products to assist in interpreting data from cDNA microarrays. CellSpace can also determine which protein-protein interactions observed in a two-hybrid screen have not previously been reported in the literature, list all the molecular

targets implicated in a disease and the drugs known to act upon them. All answers are linked back to the reports from which they were derived and new literature is automatically analysed every two weeks. Users are notified via email if an update occurs that is relevant to a particular query. A two-week free trial of Cellspace 2.0 is currently available from <http://cellspace.cellomics.com/cellspace/>.

Reference materials

LGC Promochem www.lgcpromochem.com
Combined catalogue for 2002

Reference materials assist scientists working with a broad range of analyses to improve the accuracy, traceability and repeatability of chemical measurements. This catalogue replaces LGC Promochem's previous 'Reference materials for pharmaceutical analysis' and 'Reference substances for the analysis of pharmaceutical impurities' catalogues. The new edition provides a single, comprehensive source of reference substances and pharmaceutical impurities elements.

Pharm-O-Kin

GeneScan www.genescan.com
A screen test for trial subjects

Individuals react differently to drugs, according to their genetic profiles and the metabolic activities defined in the genome. Pharmacogenetic typing provides precise predictions about the effectiveness and tolerability of a drug treatment, the optimum dose of the active substance, and the risk of adverse drug reactions. The Pharm-O-Kin Screen service is

aimed at researchers who are recruiting test subjects and patients for phase I–III clinical trials. It is based on a biochip that identifies 39 relevant polymorphisms with currently known functions coding for key enzymes of drug metabolism and transport. The results of the screen, including an interpreted report and a list of polymorphisms examined, are provided within 2–3 days of the arrival of blood samples. GeneScan expects this screening service to improve efficiency and safety in drug development trials by allowing targeted recruitment of study participants on the basis of pharmacogenetic profile.

FluoroFectin

Qbiogene www.qbiogene.com
Make tracks with these reagents

Qbiogene's new liposome-based, fluorescently labelled transfection reagents facilitate the transfer of genetic material into cells. FluoroFectin reagents were developed to enhance users' ability to visualize gene delivery and track and assay DNA uptake. Four different fluorescent labels are available. FluoroFectin Green shows the formation of the DNA–liposome complex and its distribution in the cells in the visible light range. FluoroFectin Red can combine with fluorescently labelled DNA for lipid/gene/gene product localization studies. FluoroFectin Red pH Sensitive tracks the entry of the complex into the cell and FluoroFectin UV is designed for the study of cationic liposomes with nucleic acids.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)
Amelie Pequignot (4974)
Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)
Scandinavia: Sille Opstrup (4994)

Production Manager: Billie Franklin

To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel +33 (0) 1 43 20 16 51
Fax +33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen
Tel +49 89 54 90 57 11/-2
Fax +49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakata-machi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.johnson@naturejpn.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

Fresh paths in Europe

Claude Kordon suspects that there are two important trends in European science employment — career paths are changing, and European scientists aren't fully prepared for these changes. Kordon — editor of the journal *Neuroendocrinology*, a researcher at the Necker Institute in Paris, and a former president of Euroscience, a young pan-European scientific society — reached his conclusions as a result of both anecdote and experience. And he has spent the past few years of his own career studying these European trends.

Kordon anticipates that there will be an increase in interdisciplinary interactions and a rise in short-term contracts across Europe. He says that more opportunities may arise in industry than academia, but adds that many European scientists are not aware of those opportunities — especially 'alternative' careers in areas such as law, finance and marketing — and are not prepared for them.

To complicate matters further, Kordon feels that the information currently available on the Internet about European career opportunities tends to be of limited use. "There are a certain number of databases, but usually they are local and fragmented," he says.

As a first step to providing more global and complete information, Kordon is helping Euroscience to organize a November conference on new science- and technology-based professions in Europe. The conference will act as a sort of a survey by bringing together scientists, science managers, sociologists, teachers, employers, trade unionists and representatives of young-scientists' associations. Gathering data will be the easy part. Creating new career paths and making young scientists more aware of them will be much more difficult. "We still have a long way to go," Kordon says.

Paul Smaglik

Naturejobs editor



Contents

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS